

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:00:00 ; Search time 64.99 Seconds  
(without alignments)  
61.288 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_1\_177

Perfect score: 971

Sequence: 1 MAFHGGSLITLVPMMAALL.....NTGDPHPCPENGSCVPDGP 177

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	8.4	670	1 US-08-243-542-3	Sequence 3, Appl1
2	82	8.4	670	1 US-08-477-407-3	Sequence 3, Appl1
3	82	8.4	670	1 US-08-484-355-3	Sequence 3, Appl1
4	82	8.4	769	1 US-08-243-542-4	Sequence 4, Appl1
5	82	8.4	769	1 US-08-477-407-4	Sequence 4, Appl1
6	82	8.4	769	1 US-08-484-355-4	Sequence 4, Appl1
7	79.5	8.2	321	4 US-08-915-795-9	Sequence 9, Appl1
8	79.5	8.2	358	4 US-08-915-795-8	Sequence 8, Appl1
9	78.5	8.1	806	4 US-08-945-983-2	Sequence 2, Appl1
10	76.5	7.9	2213	1 US-08-727-034-3	Sequence 3, Appl1
11	75	7.7	886	3 US-09-110-116-3	Sequence 3, Appl1
12	74	7.6	248	4 US-08-944-483-63	Sequence 63, Appl1
13	72.5	7.5	1193	2 US-08-400-159-10	Sequence 10, Appl1
14	72.5	7.5	1193	2 US-08-611-729A-10	Sequence 10, Appl1
15	72.5	7.5	1525	3 US-09-191-647-2	Sequence 2, Appl1
16	72.5	7.5	1525	4 US-09-540-245A-2	Sequence 2, Appl1
17	72.5	7.5	1525	4 US-09-540-153-2	Sequence 2, Appl1
18	71.5	7.4	225	2 US-08-557-146-12	Sequence 12, Appl1
19	71.5	7.4	225	2 US-09-154-344-12	Sequence 12, Appl1
20	71	7.3	224	4 US-08-944-483-33	Sequence 33, Appl1
21	71	7.3	225	2 US-09-027-357-4	Sequence 4, Appl1
22	71	7.3	253	2 US-08-557-146-2	Sequence 2, Appl1
23	71	7.3	253	2 US-08-824-874-3	Sequence 2, Appl1
24	71	7.3	253	2 US-09-154-344-2	Sequence 2, Appl1
25	71	7.3	253	3 US-08-930-188-2	Sequence 2, Appl1
26	71	7.3	253	4 US-09-210-084-3	Sequence 2, Appl1
27	71	7.3	253	5 PCT-US96-04294-2	Sequence 2, Appl1

28	70.5	7.3	805	3 US-08-985-526-34	Sequence 34, Appl1
29	70.5	7.3	806	2 US-08-443-861-5	Sequence 5, Appl1
30	70.5	7.3	806	4 US-08-193-829B-5	Sequence 5, Appl1
31	70.5	7.3	1367	1 US-07-813-593-4	Sequence 4, Appl1
32	70.5	7.3	1367	1 US-07-977-451-6	Sequence 6, Appl1
33	70.5	7.3	1367	1 US-07-946-507-4	Sequence 6, Appl1
34	70.5	7.3	1367	1 US-08-252-517-6	Sequence 6, Appl1
35	70.5	7.3	1367	1 US-07-906-397A-6	Sequence 6, Appl1
36	70.5	7.3	1367	1 US-08-601-891-6	Sequence 6, Appl1
37	70.5	7.3	1367	2 US-08-443-861-2	Sequence 2, Appl1
38	70.5	7.3	1367	2 US-09-021-324-6	Sequence 2, Appl1
39	70.5	7.3	1367	4 US-08-193-829B-2	Sequence 2, Appl1
40	70.5	7.3	1367	5 PCT-US92-02750-8	Sequence 8, Appl1
41	70.5	7.3	1367	5 PCT-US92-05401-6	Sequence 6, Appl1
42	70.5	7.3	1367	5 PCT-US92-09893-6	Sequence 6, Appl1
43	70	7.2	158	2 US-08-247-904B-10	Sequence 10, Appl1
44	70	7.2	158	3 US-08-767-942A-19	Sequence 19, Appl1
45	70	7.2	271	1 US-08-117-083-14	Sequence 14, Appl1

## ALIGNMENTS

RESULT 1  
US-08-243-542-3  
; Sequence 3, Application US/08243542  
; Patent No. 5552526  
GENERAL INFORMATION:  
APPLICANT: NAKAMURA, YOSUKE  
APPLICANT: EMI, MITSURU  
TITLE OF INVENTION: MDC PROTEINS AND DNAs  
TITLE OF INVENTION: ENCODING THE SAME  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS P.C.  
STREET: 2026 Rambling Road  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49008-1699  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage  
COMPUTER: IBM PC/XT/AT Compatible  
OPERATING SYSTEM: MS-DOS 5.0  
SOFTWARE: Wordperfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/243,542  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-136602  
FILING DATE: 14 MAY 1993  
APPLICATION NUMBER: JP 5-257455  
FILING DATE: 22 SEPTEMBER 1993  
APPLICATION NUMBER: JP 6-49904  
FILING DATE: 23 FEBRUARY 1994  
APPLICATION NUMBER: JP 6-73328  
FILING DATE: 12 APRIL 1994  
APPLICATION NUMBER: JP 6-84470  
FILING DATE: 22 APRIL 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Teriyence F. Chapman  
REGISTRATION NUMBER: 32 549  
REFERENCE/DOCKET NUMBER: Furuya Case 1313  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (616) 381-1156  
TELEFAX: (616) 381-5465  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 670 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein





FILING DATE: 22 APRIL 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Terryence F. Chapman  
 REGISTRATION NUMBER: 32 549  
 REFERENCE/DOCKET NUMBER: Furuya Case 1313  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (616) 381-1156  
 TELEFAX: (616) 381-5465  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 670 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 LIBRARY: human fetal brain cDNA library  
 US-08-484-355-3

Query Match 8.4%; Score 82; DB 1; Length 670;  
 Best Local Similarity 22.6%; Pred. No. 0.97;  
 Matches 47; Conservative 25; Mismatches 62; Indels 74; Gaps 14;

OY 31 LPEICT---QCPGVSQNLKVAFYCK-----TTRE-----IMLHA---RCC-- 65  
 Db 413 IAECTGDSQCPNHLKLD--GYCDHEGRCYGRCTRPDROCQYLVGHAADRFCE 470  
 OY 66 -LNKGT-----ILGLDLQNCLEDPGPNFHOAHNTVITIDLANPLKGLDANTFPGTQ 118  
 Db 471 KLVNVTGRSGCGKRGSGWQCSKQDVLGCF-----LLCVNISGAPRLDVLGDISSVTF 525  
 OY 119 LQTLILPQHVNCPEG--INAMNTITSYID-----NOIC-----OG 151  
 Db 526 YHQ--GKELDCRGHGYQLADGSDLSYEDGTACGPNMLCLDHRCLPASAFNSTCPGSG 582  
 OY 152 QKNLCNNTGDEMCPEGSCV--PDGPG 177  
 Db 583 ERRICSHHG--VCSNEGKCIQPDWTG 607

RESULT 4  
 US-08-243-542-4  
 ; Sequence 4, Application US/08243542  
 ; Patent No. 5552526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAMURA, YUSUKE  
 ; APPLICANT: EMI, MITSURU  
 ; TITLE OF INVENTION: MDC PROTEINS AND DNAs  
 ; TITLE OF INVENTION: ENCODING THE SAME  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS P.C.  
 ; STREET: 2026 Rambling Road  
 ; CITY: Kalamazoo  
 ; STATE: Michigan  
 ; COUNTRY: USA  
 ; ZIP: 49008-1699  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage  
 ; COMPUTER: IBM PC/XT/AT Compatible  
 ; OPERATING SYSTEM: MS-DOS 5.0  
 ; SOFTWARE: WordPerfect 5.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/243,542  
 ; FILING DATE:  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 5-136602  
 ; FILING DATE: 14 MAY 1993  
 ; APPLICATION NUMBER: JP 5-257455  
 ; FILING DATE: 22 SEPTEMBER 1993

APPLICATION NUMBER: JP 6-49904  
 FILING DATE: 23 FEBRUARY 1994  
 APPLICATION NUMBER: JP 6-73328  
 FILING DATE: 12 APRIL 1994  
 APPLICATION NUMBER: JP 6-84470  
 FILING DATE: 22 APRIL 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Terryence F. Chapman  
 REGISTRATION NUMBER: 32 549  
 REFERENCE/DOCKET NUMBER: Furuya Case 1313  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (616) 381-1156  
 TELEFAX: (616) 381-5465  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 769 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 LIBRARY: human fetal brain cDNA library  
 US-08-243-542-4

Query Match 8.4%; Score 82; DB 1; Length 769;  
 Best Local Similarity 22.6%; Pred. No. 1.2;  
 Matches 47; Conservative 25; Mismatches 62; Indels 74; Gaps 14;

OY 31 LPEICT---QCPGVSQNLKVAFYCK-----TTRE-----IMLHA---RCC-- 65  
 Db 512 IAECTGDSQCPNHLKLD--GYCDHEGRCYGRCTRPDROCQYLVGHAADRFCE 569  
 OY 66 -LNKGT-----ILGLDLQNCLEDPGPNFHOAHNTVITIDLANPLKGLDANTFPGTQ 118  
 Db 570 KLVNVTGRSGCGKRGSGWQCSKQDVLGCF-----LLCVNISGAPRLDVLGDISSVTF 624  
 OY 119 LQTLILPQHVNCPEG--INAMNTITSYID-----NOIC-----OG 151  
 Db 625 YHQ--GKELDCRGHGYQLADGSDLSYEDGTACGPNMLCLDHRCLPASAFNSTCPGSG 681  
 OY 152 QKNLCNNTGDEMCPEGSCV--PDGPG 177  
 Db 682 ERRICSHHG--VCSNEGKCIQPDWTG 706

RESULT 5  
 US-08-477-407-4  
 ; Sequence 4, Application US/08477407  
 ; Patent No. 5631351  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAMURA, YUSUKE  
 ; APPLICANT: EMI, MITSURU  
 ; TITLE OF INVENTION: MDC PROTEINS AND DNAs  
 ; TITLE OF INVENTION: ENCODING THE SAME  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS P.C.  
 ; STREET: 2026 Rambling Road  
 ; CITY: Kalamazoo  
 ; STATE: Michigan  
 ; COUNTRY: USA  
 ; ZIP: 49008-1699  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage  
 ; COMPUTER: IBM PC/XT/AT Compatible  
 ; OPERATING SYSTEM: MS-DOS 5.0  
 ; SOFTWARE: WordPerfect 5.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/477,407  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 530

## PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/243,542  
 FILING DATE: 13-MAY-1994  
 APPLICATION NUMBER: JP 5-136602  
 FILING DATE: 14 MAY 1993  
 APPLICATION NUMBER: JP 5-257455  
 FILING DATE: 22 SEPTEMBER 1993  
 APPLICATION NUMBER: JP 6-49904  
 FILING DATE: 23 FEBRUARY 1994  
 APPLICATION NUMBER: JP 6-73328  
 FILING DATE: 12 APRIL 1994  
 APPLICATION NUMBER: JP 6-84470  
 FILING DATE: 22 APRIL 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Terrence F. Chapman  
 REGISTRATION NUMBER: 32 549  
 REFERENCE/DOCKET NUMBER: Furuya Case 1313  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (616) 381-1156  
 TELEFAX: (616) 381-5465  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 769 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 LIBRARY: human fetal brain cDNA library  
 US-08-477-407-4

Query Match 8.4%; Score 82; DB 1; Length 769;  
 Best Local Similarity 22.6%; Pred. No. 1.2;

Matches 47; Conservative 25; Mismatches 62; Indels 74; Gaps 14;

QY 31 LPEICT-----OCPSYONLSKVAFYCK-----TTRE-----LMLHA---RCC-- 65  
 DB 512 IAECTGDSGCCPPNHLKLD--GYCDHEGRCYGRCKTRDRCQYLVGMHAADRFCE 569  
 QY 66 -LNOKGT-----ILGLDLNCSLEDPGPNFHOAHTTVIIDLANPLKGLANTFRGFTQ 118  
 DB 570 KLVNVEGTERSGCGRKSGWVCSKQDYLGF-----LLCVNISGAPRLDGLVDISSVTF 624  
 QY 119 LQTLILPQHVNCPEG--INAMNTITSYID-----NOIC-----OG 151  
 DB 625 YHQ---GKELDCRGHVLADGSDLSYEDGTACGPNMLCLDRCLPASAFNESTCPGSG 681  
 QY 152 QKNLCNNTGDEMPENGCVCV--PDGPG 177  
 DB 682 ERRICSHHG---VCSNECKICQPDWTG 706

RESULT 6  
 US-08-484-355-4  
 Sequence 4, Application US/08484355  
 Patent No. 5705341  
 GENERAL INFORMATION:  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: EMI, MITSURU  
 TITLE OF INVENTION: MDC PROTEINS AND DNAs  
 TITLE OF INVENTION: ENCODING THE SAME  
 NUMBER OF SEQUENCES: 20  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FLYNN, THIEL, BOOTHELL & TANIS P.C.  
 STREET: 2026 Rambling Road  
 CITY: Kalamazoo  
 STATE: Michigan  
 COUNTRY: USA  
 ZIP: 49008-1699  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage

COMPUTER: IBM PC/XT/AT Compatible  
 OPERATING SYSTEM: MS-DOS 5.0  
 SOFTWARE: WordPerfect 5.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/484,355  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/243,542  
 FILING DATE: 13-MAY-1994  
 APPLICATION NUMBER: JP 5-136602  
 FILING DATE: 14 MAY 1993  
 APPLICATION NUMBER: JP 5-257455  
 FILING DATE: 22 SEPTEMBER 1993  
 APPLICATION NUMBER: JP 6-49904  
 FILING DATE: 23 FEBRUARY 1994  
 APPLICATION NUMBER: JP 6-73328  
 FILING DATE: 12 APRIL 1994  
 APPLICATION NUMBER: JP 6-84470  
 FILING DATE: 22 APRIL 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Terrence F. Chapman  
 REGISTRATION NUMBER: 32 549  
 REFERENCE/DOCKET NUMBER: Furuya Case 1313  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (616) 381-1156  
 TELEFAX: (616) 381-5465  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 769 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 LIBRARY: human fetal brain cDNA library  
 US-08-484-355-4

Query Match 8.4%; Score 82; DB 1; Length 769;  
 Best Local Similarity 22.6%; Pred. No. 1.2;

Matches 47; Conservative 25; Mismatches 62; Indels 74; Gaps 14;

QY 31 LPEICT-----OCPSYONLSKVAFYCK-----TTRE-----LMLHA---RCC-- 65  
 DB 512 IAECTGDSGCCPPNHLKLD--GYCDHEGRCYGRCKTRDRCQYLVGMHAADRFCE 569  
 QY 66 -LNOKGT-----ILGLDLNCSLEDPGPNFHOAHTTVIIDLANPLKGLANTFRGFTQ 118  
 DB 570 KLVNVEGTERSGCGRKSGWVCSKQDYLGF-----LLCVNISGAPRLDGLVDISSVTF 624  
 QY 119 LQTLILPQHVNCPEG--INAMNTITSYID-----NOIC-----OG 151  
 DB 625 YHQ---GKELDCRGHVLADGSDLSYEDGTACGPNMLCLDRCLPASAFNESTCPGSG 681  
 QY 152 QKNLCNNTGDEMPENGCVCV--PDGPG 177  
 DB 682 ERRICSHHG---VCSNECKICQPDWTG 706

RESULT 7  
 US-08-915-795-9  
 Sequence 9, Application US/08915795  
 Patent No. 6235713  
 GENERAL INFORMATION:  
 APPLICANT: Marc G. ACHEN  
 APPLICANT: Andrew F. WILKS  
 APPLICANT: Steven A. STACKER  
 APPLICANT: Karl ALITALO  
 TITLE OF INVENTION: GROWTH FACTOR  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:

```

ADDRESS: Evenson, McKeown, Edwards & Lenahan P.L.L.C.
STREET: 1200 G Street, NW, Suite 700
CITY: Washington
STATE: DC
COUNTRY: United States of America
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,795
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: EVANS, Joseph D.
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 1064/42983
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
TELEX: N/A
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 321 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
TISSUE TYPE: Mouse Lung
US-08-915-795-9

Query Match      8.2%  Score 79.5;  DB 4;  Length 321;
Best Local Similarity 26.4%;  Pred. No. 0.65;
Matches 38; Conservative 14; Mismatches 49; Indels 43; Gaps 9;

QY 63 RCCLNKGFTIGLDLONCSLEDP--GPNFQAHTTVIID-----LQANPLK----- 106
DB 66 RCRLKLR-SLASMSRSASHSTRPATFTYDTETLKVIDEEMORTQCSRETCVEYASEL 124
QY 107 GDLANTFRGFTQLTLIPQVHN---CPG-----GINAMNTITSYIDNOICOGKNCNN 158
DB 125 GKTNTTF-----FKPPCVNVFRCGGCGCNBEGVMCMNTSTYSISKQLEFISVPL--- 172
QY 159 TGDEPMCP-----ENGSCVPDGP 176
DB 173 TSVPFLVPVKIANHTGCKCLPTGP 196

RESULT 8
US-08-915-795-8
Sequence 8, Application US/08915795
Patent No. 6235713
GENERAL INFORMATION:
APPLICANT: Marc G. ACHEN
APPLICANT: Andrew F. MILKS
APPLICANT: Steven A. STRACKER
APPLICANT: Kari ALITALO
TITLE OF INVENTION: GROWTH FACTOR
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan P.L.L.C.
STREET: 1200 G Street, NW, Suite 700
CITY: Washington
STATE: DC
COUNTRY: United States of America
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,795
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: EVANS, Joseph D.
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 1064/42983
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800
TELEFAX: (202) 628-8844
TELEX: N/A
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 358 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
TISSUE TYPE: Mouse Lung
US-08-915-795-8

Query Match      8.2%  Score 79.5;  DB 4;  Length 358;
Best Local Similarity 26.4%;  Pred. No. 0.76;
Matches 38; Conservative 14; Mismatches 49; Indels 43; Gaps 9;

QY 63 RCCLNKGFTIGLDLONCSLEDP--GPNFQAHTTVIID-----LQANPLK----- 106
DB 71 RCRLKLR-SLASMSRSASHSTRPATFTYDTETLKVIDEEMORTQCSRETCVEYASEL 129
QY 107 GDLANTFRGFTQLTLIPQVHN---CPG-----GINAMNTITSYIDNOICOGKNCNN 158
DB 130 GKTNTTF-----FKPPCVNVFRCGGCGCNBEGVMCMNTSTYSISKQLEFISVPL--- 177
QY 159 TGDEPMCP-----ENGSCVPDGP 176
DB 178 TSVPFLVPVKIANHTGCKCLPTGP 201

RESULT 9
US-08-945-983-2
Sequence 2, Application US/08945983
Patent No. 6225527
GENERAL INFORMATION:
APPLICANT: Thomas, Colwyn M
APPLICANT: Ballint-Kurtli, Peter J
APPLICANT: Jones, David A
APPLICANT: Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 8th Floor, 1100 No. 6225527th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,983
FILING DATE: 12-NOV-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/GB96/01155
FILING DATE: 13-MAY-1996

```

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9509575.8  
 FILING DATE: 11-MAY-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ms Mary J Wilson  
 REGISTRATION NUMBER: 32,955  
 REFERENCE/DOCKET NUMBER: 620-27  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 816-4000  
 TELEFAX: (703) 816-4100  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 806 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Lycopersicon hirsutum  
 STRAIN: Cf4  
 US-08-945-983-2

Query Match 8.1%; Score 78.5; DB 4; Length 806;  
 Best Local Similarity 26.6%; Pred. No. 3.1;  
 Matches 25; Conservative 15; Mismatches 33; Indels 21; Gaps 3;

QY 71 TILGLDLCNSLEDPGFHQHTVYIIDLANPLKGLDANLFRGFTQLOTLLPOHVC 130  
 DB 351 SLIGLDLSNTPFGKIOEF KSKLTSTVTLKQKLGPIPNLSLKNQIFILL----- 403  
 QY 131 PGGINAMNTTSTYIDNOICQ-----GQKNL 155  
 DB 404 -----SHNNISCHISAICMLKTLILLDLSNNL 432

RESULT 10  
 US-08-727-034-3  
 Sequence 3, Application US/08727034  
 Patent No. 5665872  
 GENERAL INFORMATION:  
 APPLICANT: SATTO, YASUSHI  
 APPLICANT: IMASAKI, AKIO  
 APPLICANT: ARAI, KOICHI  
 APPLICANT: YAMAZAKI, HIROYUKI  
 TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND  
 TITLE OF INVENTION: THE GENE CODING THEREFOR  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
 ADDRESSEE: P.C.  
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 CITY: ARLINGTON  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/727,034  
 FILING DATE: 08-OCT-1996  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 261440/1995  
 FILING DATE: 09-OCT-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 102451/1996  
 FILING DATE: 24-APR-1996

ATTORNEY/AGENT INFORMATION:  
 NAME: OBLON, NORMAN F.  
 REGISTRATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 80-079-0  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-413-3000  
 TELEFAX: 703-413-2220  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2213 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-727-034-3

Query Match 7.9%; Score 76.5; DB 1; Length 2213;  
 Best Local Similarity 19.2%; Pred. No. 22;  
 Matches 39; Conservative 16; Mismatches 59; Indels 89; Gaps 9;

QY 50 FYCKTRELMLHARCLNQKGTILGLDLCNSLEDPG-----PNFH-----QAHF 94  
 DB 1280 FVCKNRQQLFHSVC---DGI---QCRGSDDEDPAFAGCSRPERHKYCDERFGQCN 1333  
 QY 95 TVIIDL-----QANPLKGLDANLFRGF----- 116  
 DB 1334 GVCISLIRKCGMDCCGYSDEANENPTAPNCSRYFEGCDNGHCIPNRKCDRENDG 1393  
 QY 117 -----TQLOTLLPOHVCPGGINAMNTTSTYIDNOICQKNCNN 158  
 DB 1394 GDSDEKDCGDSHVLPSSTPAPSTCLPNRYRCGG-----ACVIDWCDGYRD-CAD 1445  
 QY 159 TGDPEMCE---NGCVPDGPG 177  
 DB 1446 GSDERACPSLPNWTATSSPSQPG 1468

RESULT 11  
 US-09-110-116-3  
 Sequence 3, Application US/09110116  
 Patent No. 6013479  
 GENERAL INFORMATION:  
 APPLICANT: Xu, Hong  
 APPLICANT: Cohan, Victoria L.  
 APPLICANT: Stuart, Susan G.  
 TITLE OF INVENTION: HUMAN EMR1-LIKE G PROTEIN COUPLED  
 TITLE OF INVENTION: RECEPTOR  
 FILE REFERENCE: PR-0550 US  
 CURRENT APPLICATION NUMBER: US/09/110,116  
 CURRENT FILING DATE: 1998-07-02  
 NUMBER OF SEQ ID NOS: 4  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 3  
 LENGTH: 886  
 TYPE: PRT  
 ORGANISM: HOMO SAPIENS  
 FEATURE:  
 OTHER INFORMATION: 784994, Genbank  
 US-09-110-116-3

Query Match 7.7%; Score 75; DB 3; Length 886;  
 Best Local Similarity 21.7%; Pred. No. 8.9;  
 Matches 35; Conservative 17; Mismatches 45; Indels 64; Gaps 9;

QY 39 PGSVQNLKVAAYCKTRELMLHARC-----CLNKGIT-----IGLDLCNSLEDPGP 88  
 DB 122 PGKPGN-----FSCVDINCLTSRVCPHSDDVNSMGYSQGVGFTSRSTED----- 172  
 QY 89 FQOAHHTVYIIDLANPLKGLDANLFRGFTQLOTLLPOHVCPGGINAMNTTSTY----- 143  
 DB 173 -----VNDCADP-----RACPEHATCN-----NIVGNATSCCN 200

OY 144 -----IDNOCOGOKNLNNTGD-PEMCPENGSCVPDGP 177  
 Db 201 PGFESSGHLSCOGKASCDEDICTEMCPINSTCT-NTPG 240

# RESULT 12

US-08-944-483-63  
 ; Sequence 63, Application US/08944483  
 ; Patent No. 6232456

## GENERAL INFORMATION:

APPLICANT: COHEN, MAURICE  
 APPLICANT: COLPITTS, TRACEY L.  
 APPLICANT: FRIEDMAN, PAULA N.  
 APPLICANT: GRANADOS, EDWARD N.  
 APPLICANT: KLASS, MICHAEL R.  
 APPLICANT: RUSSELL, JOHN C.  
 APPLICANT: STEWART, KENT D.  
 APPLICANT: STROPE, STEVEN D.  
 TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
 TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
 NUMBER OF SEQUENCES: 76  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Abbott Laboratories  
 STREET: 100 Abbott Park Road  
 CITY: Abbott Park  
 STATE: IL  
 COUNTRY: USA  
 ZIP: 60064-3500

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/944,483  
 FILING DATE:  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Becker, Cheryl L.  
 REGISTRATION NUMBER: 35,441  
 REFERENCE/DOCKET NUMBER: 6183 US. 01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 847/935-1729  
 TELEFAX: 847/938-2623  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 63:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 248 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: No. 6232456e  
 US-08-944-483-63

Query Match 7.68; Score 74; DB 4; Length 248;  
 Best Local Similarity 19.48; Pred. No. 1.9;  
 Matches 45; Conservative 25; Mismatches 40; Indels 122; Gaps 13;

OY 14 PMAALLALICGERALALPELCTOCPS-----VONL 45  
 Db 13 PWOVSLGWKTLAQRHL-----CGGSLGHQWVLTAAHCFDGLPLQDVWRTYSILNL 64  
 OY 46 SKVAFYCTT-----RELMLHARCLNKGKTIIGLDLQNCSLDPGPNFHQAHTTVIIDL 100  
 Db 65 SDIT--KDTFQSOIKETIIT-----QNYKVSSEGN-----HDIALIKL 99  
 OY 101 QAN-----PLKGLDANTFR-----GFT-----QLQTLILPQHVNPGGINMW 137

Db 100 QAPLNTTEFOKPLCLPSKSDTSTIYNWCWGTGSGEKEGELIONIL--QKVNIP----- 151  
 OY 138 NITTSYIDNOCQ-----GOKNLG--NNTGDPEMCPENG 169  
 Db 152 -----LVTNECCOKRYODYKITQRMWYACGYKSGADCKGDSGGLVYKKNIG 198

# RESULT 13

US-08-400-159-10  
 ; Sequence 10, Application US/08400159  
 ; Patent No. 5869282

## GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David  
 APPLICANT: Henrique, Domingos M.P.  
 APPLICANT: Lewis, Julian H.  
 APPLICANT: Myatt, Anna M.  
 APPLICANT: Fleming, Robert J.  
 APPLICANT: Artavanis-Tsakonas, Spyridon  
 APPLICANT: Mann, Robert S.  
 TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
 TITLE OF INVENTION: SERATE GENE AND METHODS BASED THEREON  
 NUMBER OF SEQUENCES: 20  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/400,159  
 FILING DATE: 07-MAR-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 7326-029  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-9741/8864  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1193 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-400-159-10

Query Match 7.58; Score 72.5; DB 2; Length 1193;  
 Best Local Similarity 19.53; Pred. No. 26;  
 Matches 41; Conservative 19; Mismatches 59; Indels 91; Gaps 9;

OY 29 LALPEICTOCPSVO--NLKSVAFYCKTRELML----- 60  
 Db 517 LAMDYFC-NCPEDYEGKNCSHLDHCHRTTICEVIDSCYAVASNSTPESEVRISSNWGP 575  
 OY 61 HARCLNKGKTIIGLDLQNCSLDPGPNFHQAHTTVIIDLQANPLKGLDANTFRGFTQLO 120  
 Db 576 HGKC-----KSQAGGKFTCECNKNGFTGYCHEN-----INOCESNPK-----NGT 617  
 OY 121 TLILPQHVNPBGGINANTITS-----YIDNOCOGOKNLNNTG----- 160  
 Db 618 -----CIDGVNSYKICICSDGWEGTYCETININDCSKNPCHNGTCDLVNDFECCK 668  
 OY 161 -----PEMCPENGSCVPDGP 175

Db 669 NGWKGKCHSRDSCDEATCNGSTCYDEG 698

RESULT 14  
US-08-611-729A-10  
Sequence 10, Application US/08611729A  
Patent No. 6004924

## GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myal, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
APPLICANT: Gray, Grace E.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SEPARATE GENE AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESSES:  
ADDRESSES: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,729A  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-037  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1193 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-611-729A-10

Query Match 7.5%; Score 72.5; DB 3; Length 1193;  
Best Local Similarity 19.5%; Pred. No. 26;  
Matches 41; Conservative 19; Mismatches 59; Indels 91; Gaps 9;

QY 29 LALPEICTQCPG--NLSVAFYCKTTRRLM----- 60  
DB 517 LAMDFYC-NCPDEYEGKNCSTLKHCHRTTPOEYIDSCYAVASNSPDEGVRTISSVWGP 575  
QY 61 HARCCLNKGFTLGLDNLQNCSTLEDGPNFQAHTTVIIDQANPLKGLDLANTFRGFTQLO 120  
DB 576 HGKC---KSGAGKFTCECKNKGFTGYCHEN-----INDCSNPK-----NGST--- 617  
QY 121 TLIPQHNVCPGINAMNTITS-----YIDNOCQCKNLGNNTG----- 160  
DB 618 -----CIDGVNSYKICSDGEGYCEFTNINDCSKNPCHNGTGRDLVNDFFCECK 668  
QY 161 -----DPEKCPENGSCVDPG 175  
DB 669 NGWKGKCHSRDSCDEATCNGSTCYDEG 698

RESULT 15  
US-09-191-647-2  
Sequence 2, Application US/09191647  
Patent No. 6046015

## GENERAL INFORMATION:

APPLICANT: Goodman, Corey  
APPLICANT: Kid, Thomas  
APPLICANT: Brose, Katja  
APPLICANT: Tessier-Lavigne, Marc  
TITLE OF INVENTION: Modulating Robo: Ligand Interactions  
FILE REFERENCE: B98-031-3  
CURRENT APPLICATION NUMBER: US/09/191,647  
CURRENT FILING DATE: 1998-11-13  
EARLIER APPLICATION NUMBER: 60/065,544  
EARLIER FILING DATE: 1997-11-14  
EARLIER APPLICATION NUMBER: 60/081,057  
EARLIER FILING DATE: 1998-04-07  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1525  
TYPE: PRT  
ORGANISM: human  
US-09-191-647-2

Query Match 7.5%; Score 72.5; DB 3; Length 1525;  
Best Local Similarity 27.4%; Pred. No. 37;  
Matches 40; Conservative 22; Mismatches 51; Indels 33; Gaps 9;

QY 29 LALPEICTQCPG-----SVQNLKVAFYC-KTTRRLMHARCCLNKGFTL----- 73  
DB 500 LACPEKC-RCEGTIVDCSNOKLNKIPQYTELRLN-----NNEFTYLEATGIRKTL 553  
QY 74 ----GIDLQNCSTLEDGPN-FQAHTTVIIDQANPLKGLDLANTFRGFTQLO-L-PQH 127  
DB 554 POLRKINFSNNKITDIEEGAFEGAGVNEILLTSNRLNVOHKMKKGLSLKTLMLRSNR 613  
QY 128 VNCPPG-----GINAMNTITSYIDNOI 148  
DB 614 ITCVGNDSFGLSSVRLSLY-DNOI 638

Search completed: March 28, 2002, 09:00:02  
Job time: 5199 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:16:38 ; Search time 32.18 seconds  
(without alignments)  
123.775 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_1\_177

Perfect score: 177  
Sequence: 1 MAPHPCSLTFLVPMMAALL.....NTGDPEMCPEKSCVDPDGF 177

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 132412

Minimum DB seq length: 0  
Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	3.4	8	1 US-08-633-760-3	Sequence 3, Appli
2	6	3.4	19	1 US-07-679-052A-1	Sequence 1, Appli
3	6	3.4	21	1 US-07-679-052A-9	Sequence 1, Appli
4	6	3.4	25	2 US-08-997-080-1	Sequence 9, Appli
5	6	3.4	25	2 US-08-997-362-1	Sequence 1, Appli
6	6	3.4	25	3 US-08-873-970-1	Sequence 1, Appli
7	6	3.4	25	4 US-09-093-855-1	Sequence 1, Appli
8	6	3.4	25	4 US-08-705-347A-1	Sequence 1, Appli
9	6	3.4	26	2 US-08-912-272-44	Sequence 4, Appli
10	6	3.4	27	2 US-08-968-542C-34	Sequence 1, Appli
11	6	3.4	29	4 US-09-042-353-295	Sequence 295, App
12	6	3.4	29	4 US-08-758-417A-143	Sequence 143, App
13	6	3.4	30	4 US-09-105-390-20	Sequence 20, Appl
14	5	2.8	7	1 US-08-166-930-17	Sequence 17, Appl
15	5	2.8	7	2 US-08-727-045A-17	Sequence 17, Appl
16	5	2.8	8	4 US-07-861-458C-115	Sequence 115, App
17	5	2.8	9	1 US-07-908-317-1	Sequence 1, Appli
18	5	2.8	9	2 US-08-194-981E-6	Sequence 6, Appli
19	5	2.8	9	3 US-08-159-339A-697	Sequence 697, App
20	5	2.8	9	5 PCT-US93-06171-1	Sequence 1, Appli
21	5	2.8	10	1 US-08-346-333-46	Sequence 46, Appl
22	5	2.8	10	2 US-08-764-640-59	Sequence 59, Appl
23	5	2.8	10	3 US-08-194-981E-9	Sequence 9, Appli
24	5	2.8	10	3 US-08-973-225-59	Sequence 59, Appl
25	5	2.8	10	3 US-09-244-298A-59	Sequence 59, Appl
26	5	2.8	10	4 US-09-017-631-1	Sequence 1, Appli
27	5	2.8	10	4 US-08-619-557-5	Sequence 5, Appli

28	5	2.8	10	4 US-09-516-704-59	Sequence 59, Appl
29	5	2.8	10	5 PCT-US91-07506-46	Sequence 46, Appl
30	5	2.8	11	1 US-07-699-468-2	Sequence 2, Appli
31	5	2.8	11	1 US-07-699-468-3	Sequence 3, Appli
32	5	2.8	11	1 US-07-699-468-4	Sequence 4, Appli
33	5	2.8	11	1 US-07-699-468-5	Sequence 5, Appli
34	5	2.8	11	1 US-07-699-468-6	Sequence 6, Appli
35	5	2.8	11	2 US-08-432-871C-21	Sequence 21, Appl
36	5	2.8	11	5 PCT-US93-07116-1	Sequence 1, Appli
37	5	2.8	12	1 US-08-383-743A-1	Sequence 1, Appli
38	5	2.8	12	3 US-08-808-881-1	Sequence 1, Appli
39	5	2.8	12	3 US-09-140-149-10	Sequence 10, Appl
40	5	2.8	13	2 US-08-194-981E-16	Sequence 16, Appl
41	5	2.8	13	4 US-09-258-754-184	Sequence 184, App
42	5	2.8	13	4 US-09-042-107-184	Sequence 184, App
43	5	2.8	14	1 US-07-908-317-17	Sequence 17, Appl
44	5	2.8	14	1 US-07-908-317-31	Sequence 31, Appl
45	5	2.8	14	1 US-07-908-317-34	Sequence 34, Appl

#### ALIGNMENTS

```
RESULT 1
US-08-633-760-3
; Sequence 3, Application US/08633760
; Patent No. 5804429
;
GENERAL INFORMATION:
; APPLICANT: NITWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: FUJIMURA, TAKAO
; APPLICANT: ISHII, YOSHINORI
; APPLICANT: NOGUCHI, YUJI
; TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLOK, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22206
; COMPUTER EVALUABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,760
; FILING DATE: 01-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLOK, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-929-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-633-760-3

Query Match 3.4%; Score 6; DB 1; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Caps 0;
```

OY 27 RALALP 32  
Db 3 RALALP 8

RESULT 2  
US-07-679-052A-1  
Sequence 1, Application US/07679052A  
Patent No. 5298400  
GENERAL INFORMATION:  
APPLICANT: WHITEFIELD, Peter L.  
APPLICANT: RICHARDSON, Michael A.  
APPLICANT: BUNN, Clive L.  
TITLE OF INVENTION: RECOMBINANT PRODUCT  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/679,052A  
FILING DATE: 19910506  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16786/147 CHAC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..19  
OTHER INFORMATION: /note= "Synthetic signal peptide"  
US-07-679-052A-1

Query Match 3.4%; Score 6; DB 1; Length 19;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 LLLALG 24  
Db 4 LLLALG 9

RESULT 3  
US-07-679-052A-9  
Sequence 9, Application US/07679052A  
Patent No. 5298400  
GENERAL INFORMATION:  
APPLICANT: WHITEFIELD, Peter L.  
APPLICANT: RICHARDSON, Michael A.  
APPLICANT: BUNN, Clive L.  
TITLE OF INVENTION: RECOMBINANT PRODUCT  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/679,052A  
FILING DATE: 19910506  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16786/147 CHAC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-679-052A-9

Query Match 3.4%; Score 6; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 LLLALG 24  
Db 4 LLLALG 9

RESULT 4  
US-08-997-080-1  
Sequence 1, Application US/08997080  
Patent No. 5968524  
GENERAL INFORMATION:  
APPLICANT: WATSON, JAMES D.  
APPLICANT: TAN, PAUL L.J.  
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,080  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1007



## TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563

## TELEX:

INFORMATION FOR SEQ ID NO: 1:

## SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-997-080-1

Query Match 3.4%; Score 6; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VPDGPG 177  
|||||

DB 14 VPDGPG 19

RESULT 5  
US-08-997-362-1  
Sequence 1, Application US/08997362

Patent No. 5985287

## GENERAL INFORMATION:

APPLICANT: Tan, Paul

APPLICANT: Hiyama, Jun

APPLICANT: Visser, Elizabeth

APPLICANT: Skinner, Margot

APPLICANT: Scott, Linda

APPLICANT: Prestidge, Ross

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MYCOBACTERIAL INFECTIONS

TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS

NUMBER OF SEQUENCES: 194

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/997,362

## FILING DATE:

## CLASSIFICATION:

Prior Application Number: U.S. Patent Application No. 5985287 08/873,970

Filing Date: June 12, 1997

Application Number: U.S. Patent Application No. 5985287 08/705,347

Filing Date: August 29, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet

REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1002c2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-997-362-1

Query Match 3.4%; Score 6; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VPDGPG 177  
|||||

DB 14 VPDGPG 19

RESULT 6  
US-08-873-970-1  
Sequence 1, Application US/08873970

Patent No. 6001361

## GENERAL INFORMATION:

APPLICANT: Tan, Paul

APPLICANT: Hiyama, Jun

APPLICANT: Visser, Elizabeth

APPLICANT: Skinner, Margot

APPLICANT: Scott, Linda

APPLICANT: Prestidge, Ross

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MYCOBACTERIAL INFECTIONS

TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/873,970

CLASSIFICATION: 435

Prior Application Number: 08/705,347

Filing Date: 29-AUG-1996

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet

REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1002c1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-873-970-1

Query Match 3.4%; Score 6; DB 3; Length 25;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VPDGPG 177  
|||||

DB 14 VPDGPG 19

RESULT 7

US-09-095-855-1  
Sequence 1, Application US/09095855  
Patent No. 6160093  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: Compounds and Methods for  
TREATMENT AND DIAGNOSIS OF Mycobacterial Infections  
NUMBER OF SEQUENCES: 208  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,855  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,347  
FILING DATE: 29-AUG-1996  
APPLICATION NUMBER: 08/873,970  
FILING DATE: 12-JUN-1997  
APPLICATION NUMBER: 08/997,362  
FILING DATE: 23-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-095-855-1

Query Match 3.4%; Score 6; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VPDGPG 177  
Db 14 VPDGPG 19

RESULT 8  
US-08-705-347A-1  
Sequence 1, Application US/08705347A  
Patent No. 6284255  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Hiayama, Jun  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Scott, Linda  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
DIAGNOSIS OF MYCOBACTERIAL INFECTIONS

NUMBER OF SEQUENCES: 55  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Speckman Picard PLLC  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/705,347A  
FILING DATE: 28-AUG-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269,0565  
TELEFAX: 206-269,0563  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-705-347A-1

Query Match 3.4%; Score 6; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VPDGPG 177  
Db 14 VPDGPG 19

RESULT 9  
US-08-912-272-44  
Sequence 44, Application US/08912272  
Patent No. 6093874  
GENERAL INFORMATION:  
APPLICANT: Jofuku, K. Diane  
APPLICANT: Okamuro, Jack K.  
TITLE OF INVENTION: Methods for Improving Seeds  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/912,272  
FILING DATE: 15-AUG-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/879,827  
FILING DATE: 20-JUN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/700,152

FILING DATE: 20-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-067220US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..26  
OTHER INFORMATION: /note="RAP2.7 linker region"  
US-08-912-272-44

Query Match 3.4%; Score 6; DB 3; Length 26;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 VQNLK 47  
|||||  
DB 4 VQNLK 9

RESULT 10  
US-08-968-542C-34  
Sequence 34, Application US/08968542C  
Patent No. 5981728  
GENERAL INFORMATION:  
APPLICANT: Myers, et al.  
TITLE OF INVENTION: du11 Codes For A No. 5981728el Starch  
TITLE OF INVENTION: Synthase  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McGregor & Adler, LLP  
STREET: 8011 Candle Lane  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: Microsoft Word 6.0.1 for Macintosh  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,542C  
FILING DATE: No. 5981728ember 12, 1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benjamin Aaron Adler, Ph.D., J.D.  
REGISTRATION NUMBER: 35,423  
REFERENCE/DOCKET NUMBER: D6036  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (713) 777-2321  
TELEFAX: (713) 777-6908  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acid residues  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: amino acid  
DESCRIPTION: amino acid

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
US-08-968-542C-34

Query Match 3.4%; Score 6; DB 2; Length 27;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 ILGLDL 77  
|||||  
DB 17 ILGLDL 22

RESULT 11  
US-09-042-353-295  
Sequence 295, Application US/09042353  
Patent No. 6255458  
GENERAL INFORMATION:  
APPLICANT: Lomborg, Nils  
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for  
TITLE OF INVENTION: Producing Heterologous Antibodies  
NUMBER OF SEQUENCES: 421  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3634  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/042,353  
FILING DATE: 13-MAR-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/810,279  
FILING DATE: 17-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/853,408  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/904,068  
FILING DATE: 23-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/053,131  
FILING DATE: 26-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/096,762  
FILING DATE: 22-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/155,301  
FILING DATE: 18-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,739  
FILING DATE: 03-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/165,699  
FILING DATE: 10-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/209,741  
FILING DATE: 09-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/352,322

FILED DATE: 07-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/544,404  
FILING DATE: 10-OCT-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/728,463  
FILING DATE: 10-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,739  
FILING DATE: 03-DEC-1993  
APPLICATION NUMBER: US 08/155,301  
FILING DATE: 18-NOV-1993  
APPLICATION NUMBER: US 08/096,762  
FILING DATE: 22-JUL-1993  
APPLICATION NUMBER: US 08/053,131  
FILING DATE: 26-APR-1993  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Serafini, Andrew T.  
REGISTRATION NUMBER: 41,303  
REFERENCE/DOCKET NUMBER: 014643-009030US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 143:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 amino acids  
TYPE: amino acid  
STRANDEDNESS: <unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-042-353-295

Query Match  
Best Local Similarity 3.4%; Score 6; DB 4;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 16 AAALLL 21  
Db 7 AAALLL 12

RESULT 12  
US-08-758-417A-143  
Sequence 143, Application US/08758417A  
Patent No. 6300129  
GENERAL INFORMATION:  
APPLICANT: Lomborg, Nils  
Kay, Robert M.  
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for  
Producing Heterologous Antibodies  
NUMBER OF SEQUENCES: 417  
CORRESPONDENCE ADDRESS:  
ADDRESS: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/758,417A  
FILING DATE: 02-DEC-1996  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/728,463  
FILING DATE: 10-OCT-1996  
APPLICATION NUMBER: US 08/544,404  
FILING DATE: 10-OCT-1995

APPLICATION NUMBER: US 08/352,322  
FILING DATE: 07-DEC-1994  
APPLICATION NUMBER: US 08/209,741  
FILING DATE: 09-MAR-1994  
APPLICATION NUMBER: US 08/165,699  
FILING DATE: 10-DEC-1993  
APPLICATION NUMBER: US 08/161,739  
FILING DATE: 03-DEC-1993  
APPLICATION NUMBER: US 08/155,301  
FILING DATE: 18-NOV-1993  
APPLICATION NUMBER: US 08/096,762  
FILING DATE: 22-JUL-1993  
APPLICATION NUMBER: US 08/053,131  
FILING DATE: 26-APR-1993  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Serafini, Andrew T.  
REGISTRATION NUMBER: 41,303  
REFERENCE/DOCKET NUMBER: 014643-009030US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 143:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 29 amino acids  
TYPE: amino acid  
STRANDEDNESS: <unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 143:  
US-08-758-417A-143

Query Match  
Best Local Similarity 3.4%; Score 6; DB 4;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 16 AAALLL 21  
Db 7 AAALLL 12

RESULT 13  
US-09-105-390-20  
Sequence 20, Application US/09105390  
Patent No. 6288303  
GENERAL INFORMATION:  
APPLICANT: Rodriguez, Raymond  
TITLE OF INVENTION: Rice Beta-Glucanase Enzymes  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:  
ADDRESS: Denlinger & Associates  
STREET: 350 Cambridge Ave., Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/105,390  
FILING DATE: Filed herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/050,675  
FILING DATE: 25-JUN-97  
ATTORNEY/AGENT INFORMATION:  
NAME: Petithory, Joanne R.

REGISTRATION NUMBER: P42,995  
REFERENCE/DOCKET NUMBER: 2000-0455.30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-324-0880  
TELEFAX: 650-324-0960  
TELEX:  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-09-105-390-20

Query Match 3.4%; Score 6; DB 4; Length 30;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AALLL 21  
|||||  
Db 12 AALLL 17

RESULT 14  
US-08-166-930-17  
Sequence 17, Application US/08166930  
Patent No. 559678  
GENERAL INFORMATION:  
APPLICANT: Kraus, Michael  
ATTORNEY/AGENT INFORMATION:  
TITLE OF INVENTION: Synthetic Peptides, Antibodies Against  
TITLE OF INVENTION: Them and Their Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/166,930  
FILING DATE: 15-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 42 42 736.3  
FILING DATE: 17-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 02481.1351-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-166-930-17

Query Match 2.8%; Score 5; DB 1; Length 7;

Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 CSLD 84  
|||||  
Db 1 CSLD 5

RESULT 15  
US-08-727-045A-17  
Sequence 17, Application US/08727045A  
Patent No. 5981697  
GENERAL INFORMATION:  
APPLICANT: Kraus, Michael  
ATTORNEY/AGENT INFORMATION:  
TITLE OF INVENTION: Synthetic Peptides, Antibodies Against  
TITLE OF INVENTION: Them and Their Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/727,045A  
FILING DATE: 08-OCT-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 42 42 736.3  
FILING DATE: 17-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Einaudi, Carol P.  
REGISTRATION NUMBER: 32,220  
REFERENCE/DOCKET NUMBER: 05552.1351-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-727-045A-17

Query Match 2.8%; Score 5; DB 2; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 CSLD 84  
|||||  
Db 1 CSLD 5

Search completed: March 28, 2002, 09:16:38  
Job time: 452 sec

• • •

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:00:02 ; Search time 64.99 Seconds  
(without alignments)  
18,005 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_126\_177

Sequence: 1 OHVNCPCGINAMNTITSTID.....NTGDPEMCPENSCVPPDGP 52

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2.6/prodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2.6/prodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2.6/prodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2.6/prodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2.6/prodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2.6/prodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63.5	20.6	886	3 US-09-110-116-3	Sequence 3, Appl1
2	63	20.4	2476	2 US-08-276-967-2	Sequence 2, Appl1
3	62.5	20.2	1015	1 US-08-537-210A-1	Sequence 1, Appl1
4	62.5	20.2	1015	4 US-09-113-825-1	Sequence 1, Appl1
5	62.5	20.2	2471	1 US-08-185-432-16	Sequence 16, Appl1
6	62.5	20.2	2471	1 US-08-083-590A-19	Sequence 19, Appl1
7	62.5	20.2	2471	3 US-08-532-384-19	Sequence 19, Appl1
8	61.5	19.9	2005	3 US-08-836-325-7	Sequence 7, Appl1
9	61	19.7	321	4 US-08-915-795-9	Sequence 9, Appl1
10	61	19.7	358	4 US-08-915-795-8	Sequence 8, Appl1
11	61	19.7	4654	4 US-08-476-515A-84	Sequence 84, Appl1
12	61	19.7	4655	4 US-08-652-877-84	Sequence 84, Appl1
13	61	19.7	4655	4 US-08-652-877-86	Sequence 86, Appl1
14	61	19.7	4655	4 US-08-652-877-88	Sequence 88, Appl1
15	61	19.7	4655	4 US-08-652-877-90	Sequence 90, Appl1
16	60.5	19.6	497	3 US-09-047-125-3	Sequence 3, Appl1
17	60.5	19.6	497	3 US-07-736-335E-3	Sequence 3, Appl1
18	59	19.1	441	3 US-08-985-526-3	Sequence 3, Appl1
19	57.5	18.6	1193	2 US-08-400-159-10	Sequence 10, Appl1
20	57.5	18.6	1193	3 US-08-611-729A-10	Sequence 10, Appl1
21	57	18.4	134	2 US-08-891-848-14	Sequence 14, Appl1
22	57	18.4	134	3 US-08-875-811-9	Sequence 9, Appl1
23	57	18.4	447	1 US-08-468-853-2	Sequence 2, Appl1
24	57	18.4	447	1 US-08-468-853-2	Sequence 2, Appl1
25	57	18.4	447	1 US-08-310-357-2	Sequence 2, Appl1
26	57	18.4	447	1 US-08-468-852-2	Sequence 2, Appl1
27	57	18.4	447	2 US-08-468-857-2	Sequence 2, Appl1

28	57	18.4	2213	1 US-08-727-034-3	Sequence 3, Appl1
29	55.5	18.0	673	4 US-09-075-272-5	Sequence 5, Appl1
30	55.5	18.0	1064	1 US-08-537-210A-3	Sequence 3, Appl1
31	55.5	18.0	1064	4 US-09-113-825-3	Sequence 3, Appl1
32	55.5	18.0	2523	1 US-08-185-432-18	Sequence 18, Appl1
33	54	17.5	1010	4 US-08-882-046-7	Sequence 7, Appl1
34	54	17.5	1218	2 US-08-400-159-6	Sequence 6, Appl1
35	54	17.5	1218	3 US-08-611-729A-6	Sequence 6, Appl1
36	54	17.5	1218	4 US-08-882-046-2	Sequence 2, Appl1
37	54	17.5	1218	4 US-09-214-278-7	Sequence 7, Appl1
38	54	17.5	1978	3 US-09-024-020B-3	Sequence 3, Appl1
39	54	17.5	1988	3 US-09-024-020B-4	Sequence 4, Appl1
40	54	17.5	2813	3 US-08-896-449A-2	Sequence 2, Appl1
41	54	17.5	2813	3 US-09-132-652-2	Sequence 2, Appl1
42	53	17.2	292	1 US-08-726-227-3	Sequence 3, Appl1
43	53	17.2	385	1 US-08-597-545-1	Sequence 1, Appl1
44	53	17.2	385	1 US-08-457-135-1	Sequence 1, Appl1
45	53	17.2	500	2 US-08-987-519-1	Sequence 1, Appl1

#### ALIGNMENTS

```
RESULT 1
US-09-110-116-3
: Sequence 3, Application US/09110116
: Patent No. 6013479
: GENERAL INFORMATION:
: APPLICANT: Xu, Hong
: APPLICANT: Cohen, Victoria L.
: APPLICANT: Stuart, Susan G.
: TITLE OF INVENTION: HUMAN EMRL-LIKE G PROTEIN COUPLED
: TITLE OF INVENTION: RECEPTOR
: FILE REFERENCE: PR-0550 US
: CURRENT APPLICATION NUMBER: US/09/110.116
: CURRENT FILING DATE: 1998-07-02
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 3
: LENGTH: 886
: TYPE: PR1
: ORGANISM: HOMO SAPIENS
: FEATURE:
: OTHER INFORMATION: 784994, Genbank
US-09-110-116-3

Query Match 20.6%, Score 63.5; DB 3; Length 886;
Best Local Similarity 30.0%; Pred. No. 13;
Matches 18; Conservative 7; Mismatches 22; Indels 13; Gaps 3;

QY 5 CPGGINAMNTITSTY-----IDNQCQCKNLNNTGCD-PEMCPENSCVPPDGP 52
DB 182 CPEHATCNNTGVNSFCFNCPEHSSGHLSCGKLASCEIDDECTEMCPINSTCT-NPPG 240

RESULT 2
US-08-276-967-2
: Sequence 2, Application US/08276967
: Patent No. 5851817
: GENERAL INFORMATION:
: APPLICANT: Hardy, Daniel M.
: APPLICANT: Garbers, David L.
: TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
: TITLE OF INVENTION: Sperm
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P. O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210-4433
```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/276.967  
FILING DATE: Submitted Herewith  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: UTSD:418\KIT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-787-1400  
TELEFAX: 713-789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2476 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-276-967-2

Query Match 20.4%; Score 63; DB 2; Length 2476;  
Best Local Similarity 30.4%; Pred. No. 47;  
Matches 14; Conservative 6; Mismatches 24; Indels 2; Gaps 1;

QY 5 CPGGINAWNTITSYIDNOI-CQGOKNLNNTGDEMCPEPNSGVP 48  
DB 1847 CPLDCAHSVYTSVPSCLPSCDPEGCGTGAGAPSTCEGSCICEP 1892

RESULT 3  
US-08-537-210A-1  
Sequence 1, Application US/08537210A  
Patent No. 5780300  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Fortini, Mark  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036/2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/537.210A  
FILING DATE: 29-SEP-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-027  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1015 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: hum N (Human No. 5780300ch 2)  
LOCATION: 1155...2169  
OTHER INFORMATION: Highly conserved ankyrin repeat  
US-08-537-210A-1

Query Match 20.2%; Score 62.5; DB 1; Length 1015;  
Best Local Similarity 25.3%; Pred. No. 20;  
Matches 19; Conservative 5; Mismatches 24; Indels 27; Gaps 3;

QY 4 NCPGINAW-----NTITSYIDNOICQ-----QKNLCNNTGDP 37  
DB 311 NCSSPLPCWDYINNCODELCNVECLFDNFECQGSNKTCKYDKYCAHDHFKDHNHCQGCNS 370  
QY 38 EMCPEPENG-SCVPDGP 51  
DB 371 EECGWDGIDCADCADOP 385

RESULT 4  
US-09-113-825-1  
Sequence 1, Application US/09113825  
Patent No. 6149902  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Fortini, Mark  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036/2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/113.825  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/537.210  
FILING DATE: 29-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-027  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1015 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: hum N (Human No. 6149902ch 2)



LOCATION: 1155...2169  
OTHER INFORMATION: Highly conserved ankyrin repeat  
OTHER INFORMATION: region of No. 6149902ch  
US-09-113-825-1

Query Match 20.2% Score 62.5; DB 4; Length 1015;  
Best Local Similarity 25.3%; Pred. No. 20;  
Matches 19; Conservative 5; Mismatches 24; Indels 27; Gaps 3;

QY 4 NCPGGINAW-----NTITSYIDNOICOG-----QKNLGNNTGDP 37  
DB 311 NCSSPLPCWMDYINNOCDELNTVECLFDNFECGNSKTKCKYDKYCADHFKDHCNCGNS 370  
QY 38 EMCPENG-SCVPDGP 51  
DB 371 EECGWDGLDCADDP 385

## RESULT 5

US-08-185-432-16  
Sequence 16, Application US/08185432  
Patent No. 5750652

## GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Busseau, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DEIPEX PROTEINS, NUCLEIC ACIDS, AND  
NUMBER OF SEQUENCES: 23  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530

## ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE

## INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-16

Query Match 20.2% Score 62.5; DB 1; Length 2471;  
Best Local Similarity 25.3%; Pred. No. 54;  
Matches 19; Conservative 5; Mismatches 24; Indels 27; Gaps 3;

QY 4 NCPGGINAW-----NTITSYIDNOICOG-----QKNLGNNTGDP 37  
DB 1465 NCSSPLPCWMDYINNOCDELNTVECLFDNFECGNSKTKCKYDKYCADHFKDHCNCGNS 1524

QY 38 EMCPENG-SCVPDGP 51  
DB 1525 EECGWDGLDCADDP 1539

## RESULT 6

US-08-083-590A-19  
Sequence 19, Application US/08083590A  
Patent No. 5786158

## GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/083,590A  
FILING DATE: 25-JUN-1993  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 8698864/9741  
TELEX: 66141 PENNIE

## INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-083-590A-19

Query Match 20.2% Score 62.5; DB 1; Length 2471;  
Best Local Similarity 25.3%; Pred. No. 54;  
Matches 19; Conservative 5; Mismatches 24; Indels 27; Gaps 3;

QY 4 NCPGGINAW-----NTITSYIDNOICOG-----QKNLGNNTGDP 37  
DB 1465 NCSSPLPCWMDYINNOCDELNTVECLFDNFECGNSKTKCKYDKYCADHFKDHCNCGNS 1524  
QY 38 EMCPENG-SCVPDGP 51  
DB 1525 EECGWDGLDCADDP 1539

## RESULT 7

US-08-532-384-19  
Sequence 19, Application US/08532384  
Patent No. 6083904

## GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,384  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/083,590  
FILING DATE: 25-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 869864/9741  
TELEFAX: 212 790-9090  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-532-384-19

Query Match 20.2%; Score 62.5; DB 3; Length 2471;  
Best Local Similarity 25.3%; Pred. No. 54;  
Matches 19; Conservative 5; Mismatches 24; Indels 27; Gaps 3;

OY 4 NCPGGINAM-----NTITSYIDNOCIG-----OKNICNTGDP 37  
DB 1465 NCSPPLPCMDYINNOCDELCTVCELPDFECQGNKTKCKYDKYCADHFRDNCNOCNS 1524  
OY 38 EMCPENG-SCYVDPG 51  
DB 1525 EECGMDGLDCAADP 1539

RESULT 8  
US-08-836-325-7  
Sequence 7, Application US/08836325  
Patent No. 6110672  
GENERAL INFORMATION:  
APPLICANT: Mandel, Gail  
APPLICANT: Halegoua, Simon  
APPLICANT: Borden, Laurence A.  
TITLE OF INVENTION: Peripheral Nervous System Specific  
TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,  
TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational  
TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 New York Ave., N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,325  
FILING DATE: 2-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/14251  
FILING DATE: 02-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/482,401  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/334,029  
FILING DATE: 02-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0917,0240002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2005 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-325-7

Query Match 19.9%; Score 61.5; DB 3; Length 2005;  
Best Local Similarity 33.3%; Pred. No. 56;  
Matches 16; Conservative 8; Mismatches 17; Indels 7; Gaps 3;

OY 9 INAWNTITSYIDNQ---ICQGNK-ICNNTGDEMCPENGSCYVDPG 50  
DB 311 VNMFWN-DEYIDKSHFYFLLEGNDALLCGNSSDAGOCPEGYICVKG 357

RESULT 9  
US-08-915-795-9  
Sequence 9, Application US/08915795  
Patent No. 6235713  
GENERAL INFORMATION:  
APPLICANT: Marc G. ACHEN  
APPLICANT: Andrew F. WILKS  
APPLICANT: Steven A. STRACKER  
APPLICANT: Karl ALITALO  
TITLE OF INVENTION: GROWTH FACTOR  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: EVANSON, MCKEOWN, EDWARDS & LENAHAN P.L.L.C.  
STREET: 1200 G Street, NW, Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: United States of America  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,795  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 1064/42983  
TELECOMMUNICATION INFORMATION:



Db 281 SIYKVCIGILDPCG 294

## RESULT 12

US-08-652-877-84  
; Sequence 84, Application US/08652877  
; Patent No. 6187548  
; GENERAL INFORMATION:  
; APPLICANT: Akersstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Rask, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjaln, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
; TITLE OF INVENTION: Thee of and DNA Encoding Same  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd., 3c43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.5.1  
; SOFTWARE: Word 6.0 (Patentin)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/652,877  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/15203  
; FILING DATE: 22-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/344,836  
; FILING DATE: 23-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,314  
; FILING DATE: 07-JUNE-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitzky, Martin  
; REGISTRATION NUMBER: 29,699  
; REFERENCE/DOCKET NUMBER: A1355E-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-454-3816  
; TELEFAX: 610-454-3808  
; INFORMATION FOR SEQ ID NO: 84:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4655 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-652-877-84

Query Match 19.7%; Score 61; DB 4; Length 4655;

Best Local Similarity 23.0%; Pred. No. 1.6e+02;

Matches 17; Conservative 7; Mismatches 16; Indels 34; Gaps 4;

QY 5 CPGGINAMWNTTSTYIDNOCIGOKNLNNTGDP-----MCPENGSCV 47

Db 229 CPSG-----RCIYQNMVCDGEDD-CXNDGDEDCESGPHDVHKCSPREMSCPESGRCI 280

QY 48 P-----DGRG 52

Db 281 SIYKVCIGILDPCG 294

## RESULT 13

US-08-652-877-86  
; Sequence 86, Application US/08652877  
; Patent No. 6187548  
; GENERAL INFORMATION:  
; APPLICANT: Akersstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Rask, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjaln, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
; TITLE OF INVENTION: Thee of and DNA Encoding Same  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd., 3c43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.5.1  
; SOFTWARE: Word 6.0 (Patentin)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/652,877  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/15203  
; FILING DATE: 22-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/344,836  
; FILING DATE: 23-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,314  
; FILING DATE: 07-JUNE-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitzky, Martin  
; REGISTRATION NUMBER: 29,699  
; REFERENCE/DOCKET NUMBER: A1355E-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-454-3816  
; TELEFAX: 610-454-3808  
; INFORMATION FOR SEQ ID NO: 86:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4655 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-652-877-86

Query Match 19.7%; Score 61; DB 4; Length 4655;

Best Local Similarity 23.0%; Pred. No. 1.6e+02;

Matches 17; Conservative 7; Mismatches 16; Indels 34; Gaps 4;

QY 5 CPGGINAMWNTTSTYIDNOCIGOKNLNNTGDP-----MCPENGSCV 47

Db 229 CPSG-----RCIYQNMVCDGEDD-CXNDGDEDCESGPHDVHKCSPREMSCPESGRCI 280

QY 48 P-----DGRG 52

Db 281 SIYKVCIGILDPCG 294

## RESULT 14

US-08-652-877-88  
; Sequence 88, Application US/08652877  
; Patent No. 6187548

[illegible]

Thu Mar 28 09:21:37 2002

us-09-726-348-2\_copy\_126\_177\_1.rai

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 28, 2002, 09:16:38 ; Search time 32.18 Seconds

(without alignments)  
36.363 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_126\_177

Perfect score: 52  
Sequence: 1 OHVNCPCGINAMNTITSYID.....NTGDEPCPCNGSCVDPGPG 52Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 212252 segs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 132412

Minimum DB seq length: 0  
Maximum DB seq length: 30

Post-processing: listing first 45 summaries

Database : Issued Patents.AA.\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCITUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfilest1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6	11.5	25	2	US-08-997-080-1 Sequence 1, Appli
2	6	11.5	25	2	US-08-997-362-1 Sequence 1, Appli
3	6	11.5	25	3	US-08-873-970-1 Sequence 1, Appli
4	6	11.5	25	4	US-09-095-855-1 Sequence 1, Appli
5	6	11.5	25	4	US-08-705-347A-1 Sequence 1, Appli
6	5	9.6	8	4	US-07-861-458C-115 Sequence 115, App
7	5	9.6	13	4	US-09-258-754-184 Sequence 184, App
8	5	9.6	13	4	US-09-042-107-184 Sequence 6, Appli
9	5	9.6	26	2	US-08-591-629-6 Sequence 120, App
10	4	7.7	4	1	US-08-548-540-120 Sequence 120, App
11	4	7.7	4	5	PCT-US96-09809-120 Sequence 15, Appl
12	4	7.7	6	3	US-08-526-521-15 Sequence 44, Appl
13	4	7.7	6	4	US-08-818-635-44 Sequence 83, Appl
14	4	7.7	7	1	US-08-346-333-83 Sequence 44, Appl
15	4	7.7	7	2	US-08-672-610A-44 Sequence 45, Appl
16	4	7.7	7	2	US-08-672-610A-45 Sequence 83, Appl
17	4	7.7	7	5	PCT-US91-07506-83 Sequence 26, Appl
18	4	7.7	8	1	US-08-446-206B-26 Sequence 13, Appl
19	4	7.7	8	2	US-08-807-030-13 Sequence 53, Appl
20	4	7.7	8	2	US-08-672-610A-53 Sequence 42, App
21	4	7.7	8	4	US-08-444-818-422 Sequence 423, App
22	4	7.7	8	4	US-08-444-818-423 Sequence 3, Appli
23	4	7.7	9	1	US-08-485-859-3 Sequence 10, Appl
24	4	7.7	9	1	US-08-032-846-10 Sequence 3, Appli
25	4	7.7	9	1	US-08-522-166-3 Sequence 3, Appli
26	4	7.7	9	1	US-08-488-382A-3
27	4	7.7	9	1	

28	4	7.7	9	2	US-08-480-912-3 Sequence 3, Appli
29	4	7.7	9	2	US-08-340-283-29 Sequence 29, Appl
30	4	7.7	9	2	US-08-340-283-42 Sequence 42, Appl
31	4	7.7	9	2	US-08-340-283-126 Sequence 126, App
32	4	7.7	9	2	US-08-340-283-170 Sequence 170, App
33	4	7.7	9	3	US-08-742-243-1 Sequence 1, Appli
34	4	7.7	9	3	US-08-159-339A-127 Sequence 127, App
35	4	7.7	9	3	US-08-159-339A-128 Sequence 128, App
36	4	7.7	9	4	US-09-258-754-365 Sequence 365, App
37	4	7.7	9	4	US-09-042-107-365 Sequence 6, Appli
38	4	7.7	9	5	PCT-US95-11127-19 Sequence 6, Appli
39	4	7.7	10	1	US-08-462-949-6 Sequence 69, Appl
40	4	7.7	10	1	US-08-250-789A-69 Sequence 62, Appl
41	4	7.7	10	1	US-08-250-789A-92 Sequence 6, Appli
42	4	7.7	10	1	US-08-023-764B-6 Sequence 28, Appl
43	4	7.7	10	1	US-08-214-650-28 Sequence 11, Appl
44	4	7.7	10	1	US-08-464-235-11 Sequence 23, Appl
45	4	7.7	10	2	US-08-724-548-23

## ALIGNMENTS

RESULT 1  
US-08-997-080-1  
; Sequence 1, Application US/08997080  
; Patient No. 5968524  
GENERAL INFORMATION:  
; APPLICANT: WATSON, JAMES D.  
; APPLICANT: TAN, PAUL L.  
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-  
; NUMBER OF SEQUENCES: 194  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Law Offices of Ann W. Speckman  
; STREET: 2601 Elliott Avenue, Suite 4185  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98121  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS for Windows Version 2.0  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/997,080  
; FILING DATE:  
CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
; NAME: Sleath, Janet  
; REGISTRATION NUMBER: 37,007  
; REFERENCE/DOCKET NUMBER: 11000.1007  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-269-0565  
; TELEFAX: 206-269-0563  
; TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-997-080-1  
Query Match 11.5%; Score 6; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPDGPG 52  
Db 14 VPDGPG 19

## RESULT 2

US-08-997-362-1  
Sequence 1, Application US/08997362  
Patent No. 5985287  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Hiyma, Jun  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Scott, Linda  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,362  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970  
FILING DATE: June 12, 1997  
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347  
FILING DATE: August 29, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002c2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-997-362-1

Query Match 11.5%; Score 6; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPDGPG 52  
Db 14 VPDGPG 19

RESULT 3  
US-08-873-970-1  
Sequence 1, Application US/08873970  
Patent No. 6001361  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul

APPLICANT: Hiyma, Jun  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Scott, Linda  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/873,970  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/705,347  
FILING DATE: 29-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1002c1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-873-970-1

Query Match 11.5%; Score 6; DB 3; Length 25;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPDGPG 52  
Db 14 VPDGPG 19

RESULT 4  
US-09-095-855-1  
Sequence 1, Application US/09095855  
Patent No. 6160093  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: Compounds and Methods for  
Treatment and Diagnosis of Mycobacterial Infections  
NUMBER OF SEQUENCES: 208  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121



COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PASTESEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/095,855  
FILING DATE:  
CLASSIFICATION:  
Prior APPLICATION DATA:  
APPLICATION NUMBER: 08/705,347  
FILING DATE: 29-AUG-1996  
APPLICATION NUMBER: 08/873,970  
FILING DATE: 12-JUN-1997  
APPLICATION NUMBER: 08/997,362  
FILING DATE: 23-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000,1002c3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-095-855-1

Query Match  
Best Local Similarity 11.5%; Score 6; DB 4; Length 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPDGPG 52  
DB 14 VPDGPG 19

RESULT 5  
US-08-705-347A-1  
Sequence 1, Application US/08705347A  
Patent No. 6284255  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Miyama, Jun  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Scott, Linda  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
NUMBER OF SEQUENCES: 55  
TITLE OF INVENTION: DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Speckman Picard PLLC  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/705,347A  
FILING DATE: 28-AUG-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000,1002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206,269,0565  
TELEFAX: 206,269,0563  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-705-347A-1

Query Match  
Best Local Similarity 11.5%; Score 6; DB 4; Length 25;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPDGPG 52  
DB 14 VPDGPG 19

RESULT 6  
US-07-861-458C-115  
Sequence 115, Application US/07861458C  
Patent No. 6232061  
GENERAL INFORMATION:  
APPLICANT: Johnson, Mark Andrew  
APPLICANT: Johnson, Carl D.  
TITLE OF INVENTION: HOMOLOGY CLONING  
NUMBER OF SEQUENCES: 142  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 50Z or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/861,458C  
FILING DATE: 04/01/92  
CLASSIFICATION: 435  
Prior APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 04585/014001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 115:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8  
TYPE: amino acid  
TOPOLOGY: linear  
US-07-861-458C-115

Query Match  
Best Local Similarity 9.6%; Score 5; DB 4; Length 8;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 CVPDG 50  
 Db 4 CVPDG 8

RESULT 7  
 US-09-258-754-184

; Sequence 184, Application US/09258754  
 ; Patent No. 6174687

; GENERAL INFORMATION:  
 ; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata  
 ; APPLICANT: Rajotte, Daniel

; TITLE OF INVENTION: Methods of Identifying Lung Homing Molecules Using  
 ; FILE REFERENCE: P-LJ 3443

; CURRENT APPLICATION NUMBER: US/09/258,754  
 ; EARLIER APPLICATION NUMBER: 09/042,107

; EARLIER FILING DATE: 1998-03-13  
 ; NUMBER OF SEQ ID NOS: 452

; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 184

; LENGTH: 13  
 ; TYPE: PRT

; ORGANISM: Artificial Sequence  
 ; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 US-09-258-754-184

Query Match

Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 GSCVP 48  
 Db 7 GSCVP 11

RESULT 8  
 US-09-042-107-184

; Sequence 184, Application US/09042107  
 ; Patent No. 6232287

; GENERAL INFORMATION:  
 ; APPLICANT: Ruoslahti, Erkki

; APPLICANT: Pasqualini, Renata  
 ; TITLE OF INVENTION: Molecules that Home to Various Selected Organs or

; FILE REFERENCE: P-LJ 2892  
 ; CURRENT APPLICATION NUMBER: US/09/042,107

; CURRENT FILING DATE: 1998-03-13  
 ; NUMBER OF SEQ ID NOS: 436

; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 184

; LENGTH: 13  
 ; TYPE: PRT

; ORGANISM: Artificial Sequence  
 ; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 US-09-042-107-184

Query Match

Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 GSCVP 48  
 Db 7 GSCVP 11

RESULT 9

US-08-591-629-6

; Sequence 6, Application US/08591629  
 ; Patent No. 593808

; GENERAL INFORMATION:  
 ; APPLICANT: MELCHERS, Leo Sjoerd

; APPLICANT: APOTHEKER-DE GROOT, Marion  
 ; APPLICANT: BOL, John Ferdinand

; APPLICANT: CORNELISSEN, Bernardus Johannes Clemens  
 ; APPLICANT: LINTHORST, Hubertus Josephus Maria

; APPLICANT: FONSTEIN, Anne Sijne  
 ; TITLE OF INVENTION: Plant chitinases, DNA coding therefor and  
 ; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Ladas & Parry  
 ; STREET: 26 West 61st Street  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: USA  
 ; ZIP: 10023-7604

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
 ; COMPUTER: IBM PC 4.86 SX 50 Mhz  
 ; OPERATING SYSTEM: DOS 6.20  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/591,629  
 ; FILING DATE: 15-FEB-96  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/EP94/02761  
 ; FILING DATE: 17-AUG-94  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: EP 93202425.0  
 ; FILING DATE: 17-AUG-93  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MASS, CLIFFORD J.  
 ; REGISTRATION NUMBER: 30,086  
 ; REFERENCE/DOCKET NUMBER: U-010627-0  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 708-1800  
 ; TELEFAX: (212) 246-8959  
 ; TELEX: 233288  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 26 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; HYPOTHEICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Nicotiana tabacum  
 ; DEVELOPMENTAL STAGE: TMV-induced  
 ; TISSUE TYPE: leaf  
 ; FEATURE:  
 ; NAME/KEY: peptide  
 ; LOCATION: 12  
 ; OTHER INFORMATION: /label= x  
 ; OTHER INFORMATION: /note= "both Val and Ile occur"  
 US-08-591-629-6

Query Match

Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GINAW 12  
 Db 17 GINAW 21

RESULT 10  
US-08-548-540-120  
; Sequence 120, Application US/08548540  
; Patent No. 5733731  
; GENERAL INFORMATION:  
; APPLICANT: Schatz, Peter J.  
; APPLICANT: Cull, Millard G.  
; APPLICANT: Miller, Jeff F.  
; APPLICANT: Stemmer, William P.C.  
; APPLICANT: Gates, Christian M.  
; TITLE OF INVENTION: Peptide Library and Screening Method  
; NUMBER OF SEQUENCES: 162  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: William M. Smith  
; STREET: One Market Plaza, Stewart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/548,540  
; FILING DATE: 26-OCT-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/290,641  
; FILING DATE: 15-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/963,321  
; FILING DATE: 15-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M.  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 16528J-001240US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 120:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-548-540-120

Query Match 7.7%; Score 4; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 SCVP 48  
Db 1 SCVP 4

RESULT 11  
PCT-US96-09809-120  
; Sequence 120, Application PC/TUS9609809  
; GENERAL INFORMATION:  
; APPLICANT: Schatz, Peter J.  
; APPLICANT: Cull, Millard G.  
; APPLICANT: Miller, Jeff F.  
; APPLICANT: Stemmer, William P.C.  
; APPLICANT: Gates, Christian M.  
; TITLE OF INVENTION: Peptide Library and Screening Method  
; NUMBER OF SEQUENCES: 162  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: William M. Smith

STREET: One Market Plaza, Stewart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/09809  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/548,540  
; FILING DATE: 26-OCT-1995  
; APPLICATION NUMBER: US 08/290,641  
; FILING DATE: 15-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/963,321  
; FILING DATE: 15-OCT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M.  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 16528J-001240US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-326-2400  
; TELEFAX: 415-326-2422  
; INFORMATION FOR SEQ ID NO: 120:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; PCT-US96-09809-120

Query Match 7.7%; Score 4; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 SCVP 48  
Db 1 SCVP 4

RESULT 12  
US-08-140-188-15  
; Sequence 15, Application US/08140188  
; Patent No. 5538884  
; GENERAL INFORMATION:  
; APPLICANT: Dorteich, Kurt  
; APPLICANT: Dalboge, Henrik  
; APPLICANT: Mikkelson, Jan M.  
; APPLICANT: Christensen, Flemming M.  
; APPLICANT: Halikier, Torben  
; TITLE OF INVENTION: RHAMNOLACTURONASE, CORRESPONDING DNA  
; TITLE OF INVENTION: SEQUENCE, RHAMNOLACTURONASE CONTAINING ENZYME  
; PREPARATION AND USE OF THE ENZYME PREPARATION  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5538884disk of No. 5538884th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10174-6201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/140,188  
FILING DATE: 02-NOV-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: BE 916100039.9  
FILING DATE: 02-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK92/00143  
FILING DATE: 01-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Agtis, Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3542.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: protein  
TOPOLOGY: linear  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Irpex lacteus  
STRAIN: atcc 20157  
US-08-140-188-15

Query Match 7.7%; Score 4; DB 1; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PDGP 51  
||||  
DB 2 PDGP 5

RESULT 13  
US-08-526-521-15  
Sequence 15, Application US/08526521  
Patent No. 6001627  
GENERAL INFORMATION:  
APPLICANT: Dorrelich, Kurt  
APPLICANT: Dalboge, Henrik  
APPLICANT: Mikkelsen, Jan M.  
APPLICANT: Christensen, Flemming M.  
APPLICANT: Halkier, Torben  
TITLE OF INVENTION: RHANNOGALACTURONASE, CORRESPONDING DNA  
TITLE OF INVENTION: SEQUENCE, RHANNOGALACTURONASE CONTAINING ENZYME  
TITLE OF INVENTION: PREPARATION AND USE OF THE ENZYME PREPARATION  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 60016270 No. 6001627disk of No. 6001627th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/526,521  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/140,188  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: BE 916100039.9  
FILING DATE: 02-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK92/00143  
FILING DATE: 01-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Agtis, Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3542.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
MOLECULE TYPE: protein  
TOPOLOGY: linear  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Irpex lacteus  
STRAIN: atcc 20157  
US-08-526-521-15

Query Match 7.7%; Score 4; DB 3; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PDGP 51  
||||  
DB 2 PDGP 5

RESULT 14  
US-08-818-655-44  
Sequence 44, Application US/08818655  
Patent No. 6258557  
GENERAL INFORMATION:  
APPLICANT: Lee, Mu-En  
APPLICANT: Haber, Edgar  
APPLICANT: Jain, Mukesh  
APPLICANT: Yel, Shaw-Fang  
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/818,655  
FILING DATE: 14-MAR-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/616,368  
FILING DATE: 15-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/030001

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-542-5070  
TELEFAX: 617-542-8906  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-818-655-44

Query Match 7.7%; Score 4; DB 4; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 NNTG 35  
DB 3 NNTG 6

RESULT 15  
US-08-346-333-83  
Sequence 83, Application US/08346333  
Patent No. 5677153  
GENERAL INFORMATION:  
APPLICANT: Botstein, David  
APPLICANT: Palzkill, Timothy  
TITLE OF INVENTION: Methods for modifying DNA and for  
TITLE OF INVENTION: detecting effects of such modification on interaction of  
TITLE OF INVENTION: encoded modified polypeptides with target substrates.  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard F. Trecartlin  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/346,333  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/039,501  
FILING DATE:  
APPLICATION NUMBER: US 07/602,158  
FILING DATE: 22-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartlin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-53469/RFT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
INFORMATION FOR SEQ ID NO: 83:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-346-333-83

Query Match 7.7%; Score 4; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 ENGS 45  
DB 4 ENGS 7

Search completed: March 28, 2002, 09:16:39  
Job time: 453 sec

Thu Mar 28 09:21:35 2002

us-09-726-348-2\_copy\_126\_177.rat

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 07:33:23 ; Search time 64.99 Seconds  
(without alignments)  
79.293 Million cell updates/sec

Title: US-09-726-348-2

Perfect score: 1243  
Sequence: 1 MAPHGGSLTLPMAALL.....TLLSYILLMATORRKAATS 229

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 21252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 21252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	8.4	2703	1 US-08-185-432-19	Sequence 19, Appl
2	102.5	8.2	1404	2 US-08-400-159-2	Sequence 2, Appl
3	102.5	8.2	1404	3 US-08-611-729A-2	Sequence 2, Appl
4	101.5	8.2	1139	4 US-08-537-210A-4	Sequence 4, Appl
5	101.5	8.2	1139	4 US-09-113-825-4	Sequence 4, Appl
6	99	8.0	670	1 US-08-243-542-3	Sequence 3, Appl
7	99	8.0	670	1 US-08-477-407-3	Sequence 3, Appl
8	99	8.0	670	1 US-08-464-355-3	Sequence 3, Appl
9	99	8.0	769	1 US-08-243-542-4	Sequence 4, Appl
10	99	8.0	769	1 US-08-477-407-4	Sequence 4, Appl
11	99	8.0	769	1 US-08-484-355-4	Sequence 4, Appl
12	96.5	7.8	2199	5 PCT-US95-1168A-2	Sequence 2, Appl
13	96.5	7.8	2471	1 US-08-185-432-16	Sequence 16, Appl
14	96.5	7.8	2471	1 US-08-083-590A-19	Sequence 19, Appl
15	96.5	7.8	2471	3 US-08-532-384-19	Sequence 3, Appl
16	96	7.7	886	3 US-09-110-116-3	Sequence 10, Appl
17	96	7.7	1193	2 US-08-400-159-10	Sequence 10, Appl
18	96	7.7	1193	3 US-08-611-729A-10	Sequence 10, Appl
19	92.5	7.4	1010	4 US-08-882-046-7	Sequence 7, Appl
20	92.5	7.4	1218	2 US-08-400-159-6	Sequence 6, Appl
21	92.5	7.4	1218	3 US-08-611-729A-6	Sequence 6, Appl
22	92.5	7.4	1218	4 US-08-882-046-2	Sequence 2, Appl
23	92	7.4	2556	1 US-08-185-432-17	Sequence 17, Appl
24	92	7.4	2556	1 US-08-083-590A-20	Sequence 20, Appl
25	92	7.4	2556	1 US-08-532-384-20	Sequence 20, Appl
26	91.5	7.3	1394	6 5177197-30	Patent No. 5177197
27	91	7.4	383	1 US-08-597-545-2	Sequence 2, Appl

28	91	7.3	383	1 US-08-457-135-2	Sequence 2, Appl
29	89	7.2	385	1 US-08-597-545-1	Sequence 1, Appl
30	89	7.2	385	1 US-08-457-135-1	Sequence 1, Appl
31	88.5	7.1	1218	4 US-09-214-278-7	Sequence 7, Appl
32	88.5	7.1	1480	3 US-09-191-647-7	Sequence 7, Appl
33	88.5	7.1	1480	4 US-09-540-245A-7	Sequence 7, Appl
34	88.5	7.1	1480	4 US-09-540-153-7	Sequence 7, Appl
35	88.5	7.1	1480	5 PCT-US91-09055-2	Sequence 2, Appl
36	87.5	7.0	1219	4 US-08-882-046-5	Sequence 5, Appl
37	87	7.0	1833	3 US-08-479-722B-2	Sequence 2, Appl
38	87	7.0	1833	5 PCT-US95-02251-18	Sequence 18, Appl
39	87	7.0	2523	1 US-08-185-432-18	Sequence 18, Appl
40	86.5	7.0	77	1 US-08-264-534-1	Sequence 1, Appl
41	86.5	7.0	77	1 US-08-083-590A-14	Sequence 14, Appl
42	86.5	7.0	77	1 US-08-465-500-1	Sequence 1, Appl
43	86.5	7.0	77	2 US-08-346-126-1	Sequence 1, Appl
44	86.5	7.0	77	2 US-08-346-128-1	Sequence 1, Appl
45	86.5	7.0	77	3 US-08-532-384-14	Sequence 14, Appl

#### ALIGNMENTS

RESULT 1  
US-08-185-432-19  
; Sequence 19, Application US/08185432  
; Patent No. 5750652  
GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Bussseau, Isabelle  
; APPLICANT: Diederich, Robert J.  
; APPLICANT: Xu, Tian  
; APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DELTEx PROTEINS, NUCLEIC ACIDS, AND  
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/185,432  
; FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2703 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-185-432-19

Query Match 8.4%; Score 105; DB 1; Length 2703;  
Best Local Similarity 24.0%; Pred. No. 0.1;  
Matches 47; Conservative 17; Mismatches 59; Indels 68; Gaps 10;

```

QY 27 RALALPEICTQCPGVSQNLKVAF-----YCK-----TTRMLHARC 63
DB 1147 KGLSLRQLCNNGTCKDYNSHVCYCSGAGSYCKEIDEQSPCCONGGTCRLIGAYE 1206
QY 64 CLNKGKGTLLGLDQNCLEDPGFNFHQHTTVIIDLQANPLK--GDLANTFRGFTQLOT 121
DB 1207 CQCRG-----GGQGNCELN-----IDCAPNCCQNGGTCGHDVNMFS----- 1244
QY 122 LILPQVHNCPPGINAMNTTSTYIDNQCQGNKLCNNTGDEMPCEPNSCVPDPGGLQC 181
DB 1245 -----CCCPG-----TWGI---ICEINKDCK---PGACHNNGSCI-DHVGFGEC 1283
QY 182 VCADGEHGYKC 192
DB 1284 VCPGFGVAGRC 1294

RESULT 2
US-08-400-159-2
; Sequence 2, Application US/08400159
; Patent No. 5869282
; GENERAL INFORMATION:
; APPLICANT: Ish-Horowitz, David
; APPLICANT: Henrique, Domingos M.P.
; APPLICANT: Lewis, Julian H.
; APPLICANT: Myat, Anna M.
; APPLICANT: Fleming, Robert J.
; APPLICANT: Aravanis-Tsakonas, Spyridon
; APPLICANT: Mann, Robert S.
; APPLICANT: Gray, Grace E.
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400.159
; FILING DATE: 07-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-400-159-2

Query Match 8.2%; Score 102.5; DB 2; Length 1404;
Best Local Similarity 22.7%; Pred. No. 0.076;
Matches 40; Conservative 24; Mismatches 65; Indels 47; Gaps 11;
QY 37 QCPGVSQNLKVAFYCKTTRMLHARC-----CLNKGKGT-----LGLDQNC 81
DB 236 QCAVITYNTTCTTF-CRPRDDQFGHYACGSEQKLCNLNGQGVNCEEAICKAGCDPVHKG 294

```

```

QY 82 LEDPGNFHQHTTVIIDLQANP-LKGDLANTRFGFTQLOTLLP--QHVNCGGINAMN 138
DB 295 CDRPG-----EEECRPGWRGPICN-----ECWYIPCKKSGCNG--SAWK 332
QY 139 TI--TSYIDNQCQGNKLCNNTGDEMPCEPNSCVPDPGGLQCVCADGFGGYKC 192
DB 333 CVCDDTNM--GGIILDDQDLNFC---CTHEPKHGSGTCENTAPDXRYRCACGLSGEGC 384

RESULT 3
US-08-611-729A-2
; Sequence 2, Application US/08611729A
; Patent No. 6004924
; GENERAL INFORMATION:
; APPLICANT: Ish-Horowitz, David
; APPLICANT: Henrique, Domingos M.P.
; APPLICANT: Lewis, Julian H.
; APPLICANT: Myat, Anna M.
; APPLICANT: Fleming, Robert J.
; APPLICANT: Aravanis-Tsakonas, Spyridon
; APPLICANT: Mann, Robert S.
; APPLICANT: Gray, Grace E.
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,729A
; FILING DATE: 06-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-037
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-611-729A-2

Query Match 8.2%; Score 102.5; DB 3; Length 1404;
Best Local Similarity 22.7%; Pred. No. 0.076;
Matches 40; Conservative 24; Mismatches 65; Indels 47; Gaps 11;
QY 37 QCPGVSQNLKVAFYCKTTRMLHARC-----CLNKGKGT-----LGLDQNC 81
DB 236 QCAVITYNTTCTTF-CRPRDDQFGHYACGSEQKLCNLNGQGVNCEEAICKAGCDPVHKG 294
QY 82 LEDPGNFHQHTTVIIDLQANP-LKGDLANTRFGFTQLOTLLP--QHVNCGGINAMN 138
DB 295 CDRPG-----EEECRPGWRGPICN-----ECWYIPCKKSGCNG--SAWK 332
QY 139 TI--TSYIDNQCQGNKLCNNTGDEMPCEPNSCVPDPGGLQCVCADGFGGYKC 192

```



Db 333 CVCDTNW-GGILCDDLNFC--GTHPECKHGTCENTAPDKYRCCTCAEGLSGEBC 384

## RESULT 4

US-08-537-210A-4

; Sequence 4, Application US/08537210A  
; Patent No. 5780300

## ; GENERAL INFORMATION:

; APPLICANT: Artaanis-Tsakonas, Spyridon

; APPLICANT: Fortini, Mark

; APPLICANT: Matsuno, Kenji

; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY

; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036/2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/537,210A

; FILING DATE: 29-SEP-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Mistock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-027

; TELEPHONE: 212-790-9090

; TELEFAX: 212-869-8864

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1139 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; FEATURE:

; NAME/KEY: Dros N

; LOCATION: 1189...2327

; OTHER INFORMATION: Highly conserved ankyrin repeat

; OTHER INFORMATION: region of No. 5780300ch

US-08-537-210A-4

## Query Match

Best Local Similarity 8.2%; Score 101.5; DB 1; Length 1139;

Matches 40; Conservative 11; Mismatches 45; Indels 45; Gaps 8;

OY 54 TTRELMHARCCLNOKGTITGLDNLCSLEDPGNFHAHTTYIIDQANPLK--GDLAN 111

Db 9 TCRDLIGAYECQCRQ-----GFGQNCLELN-----IDDCAPNCPQNGSTGCHD 50

OY 112 TFRGFTQLQTLILPQHVNCPGGINAMWTTTSYIDNOCQGNKLCNNTGDPNCPENGSC 171

Db 51 RVNMF-----CSCPPG-----TMGI---ICEINKDDCK----PGACHNNGSC 86

OY 172 VPDGPGILQCYCAGDFHGYKC 192

Db 87 I-DRVGGFECVCPGFGVAGRC 106

## RESULT 5

US-09-113-825-4

; Sequence 4, Application US/09113825

; Patent No. 6149902

## ; GENERAL INFORMATION:

; APPLICANT: Artaanis-Tsakonas, Spyridon

; APPLICANT: Fortini, Mark

; APPLICANT: Matsuno, Kenji

; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY

; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036/2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/113,825

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/537,210

; FILING DATE: 29-SEP-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Mistock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-027

; TELEPHONE: 212-790-9090

; TELEFAX: 212-869-8864

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1139 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; FEATURE:

; NAME/KEY: Dros N

; LOCATION: 1189...2327

; OTHER INFORMATION: Highly conserved ankyrin repeat

; OTHER INFORMATION: region of No. 6149902ch

US-09-113-825-4

## Query Match

Best Local Similarity 8.2%; Score 101.5; DB 4; Length 1139;

Matches 40; Conservative 11; Mismatches 45; Indels 45; Gaps 8;

OY 54 TTRELMHARCCLNOKGTITGLDNLCSLEDPGNFHAHTTYIIDQANPLK--GDLAN 111

Db 9 TCRDLIGAYECQCRQ-----GFGQNCLELN-----IDDCAPNCPQNGSTGCHD 50

OY 112 TFRGFTQLQTLILPQHVNCPGGINAMWTTTSYIDNOCQGNKLCNNTGDPNCPENGSC 171

Db 51 RVNMF-----CSCPPG-----TMGI---ICEINKDDCK----PGACHNNGSC 86

OY 172 VPDGPGILQCYCAGDFHGYKC 192

Db 87 I-DRVGGFECVCPGFGVAGRC 106

## RESULT 6

US-08-243-542-3

; Sequence 3, Application US/08243542

; Patent No. 5552526

## ; GENERAL INFORMATION:

; APPLICANT: NAKAMURA, YUSUKE

APPLICANT: EMI, MITSURU  
TITLE OF INVENTION: MDC PROTEINS AND DNAs  
TITLE OF INVENTION: ENCODING THE SAME  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLYNN, THEL, BOUTELL & TANIS P.C.  
STREET: 2026 Rambling Road  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49008-1699  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage  
COMPUTER: IBM PC/XT/AT Compatible  
OPERATING SYSTEM: MS-DOS 5.0  
SOFTWARE: WordPerfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/243,542  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-136602  
FILING DATE: 14 MAY 1993  
APPLICATION NUMBER: JP 5-257455  
FILING DATE: 22 SEPTEMBER 1993  
APPLICATION NUMBER: JP 6-49904  
FILING DATE: 23 FEBRUARY 1994  
APPLICATION NUMBER: JP 6-73328  
FILING DATE: 12 APRIL 1994  
APPLICATION NUMBER: JP 6-84470  
FILING DATE: 22 APRIL 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Terryence F. Chapman  
REGISTRATION NUMBER: 32 549  
REFERENCE/DOCKET NUMBER: Furuya Case 1313  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (616) 381-1156  
TELEFAX: (616) 381-5465  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 670 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
LIBRARY: human fetal brain cDNA library  
US-08-243-542-3

Query Match 8.0%; Score 99; DB 1; Length 670;  
Best Local Similarity 21.4%; Pred. No. 0.063;  
Matches 55; Conservative 32; Mismatches 90; Indels 80; Gaps 14;  
QY 31 LPEICT-----QCPGSVQNLKVAFYCK-----TTRE-----LMLHA---RCC-- 65  
DB 413 IAEFTGDSGSCGRKSGWVQCSKQDVLGCF-----LLCVNISGAPRLGDLVGISSVTF 525  
QY 66 -LNOKGT-----ILGLDQNLCSLEDPGPNFHQAHTTVIIDLQANPLKGLDLANTFRGFTQ 118  
DB 471 KLVNVEGTERGSCGRKSGWVQCSKQDVLGCF-----LLCVNISGAPRLGDLVGISSVTF 525  
QY 119 IOTLLPQHVNCPCG--INAMNTTISYD-----NOICGQKRL-----CNNTG 160  
DB 526 YHO--GKELDRCGHVOLDABSDSLVEDGTACGPNMLCDHRLCPASAFNEFTCPESG 582  
QY 161 DEPMCPKSGCVPRDGLLOCVCADGPHGYC-----MROGSFSLMFF 204  
DB 583 ERRTCSHHGVCNKG-----KICQPDWTGKDCSTHNPFTSPPIGETERTYKGPSTNII 638  
QY 205 GILGATTLVSILLMAT 221

DB 639 GS1AGAVIVAATVLTGST 655  
RESULT 7  
US-08-477-407-3  
Sequence 3, Application US/08477407  
Patent No. 5631351  
GENERAL INFORMATION:  
APPLICANT: NAKAMURA, YUSUKE  
APPLICANT: EMI, MITSURU  
TITLE OF INVENTION: MDC PROTEINS AND DNAs  
TITLE OF INVENTION: ENCODING THE SAME  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLYNN, THEL, BOUTELL & TANIS P.C.  
STREET: 2026 Rambling Road  
CITY: Kalamazoo  
STATE: Michigan  
COUNTRY: USA  
ZIP: 49008-1699  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage  
COMPUTER: IBM PC/XT/AT Compatible  
OPERATING SYSTEM: MS-DOS 5.0  
SOFTWARE: WordPerfect 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,407  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/243,542  
FILING DATE: 13-MAY-1994  
APPLICATION NUMBER: JP 5-136602  
FILING DATE: 14 MAY 1993  
APPLICATION NUMBER: JP 3-257455  
FILING DATE: 22 SEPTEMBER 1993  
APPLICATION NUMBER: JP 6-49904  
FILING DATE: 23 FEBRUARY 1994  
APPLICATION NUMBER: JP 6-73328  
FILING DATE: 12 APRIL 1994  
APPLICATION NUMBER: JP 6-84470  
FILING DATE: 22 APRIL 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Terryence F. Chapman  
REGISTRATION NUMBER: 32 549  
REFERENCE/DOCKET NUMBER: Furuya Case 1313  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (616) 381-1156  
TELEFAX: (616) 381-5465  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 670 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
LIBRARY: human fetal brain cDNA library  
US-08-477-407-3  
Query Match 8.0%; Score 99; DB 1; Length 670;  
Best Local Similarity 21.4%; Pred. No. 0.063;  
Matches 55; Conservative 32; Mismatches 90; Indels 80; Gaps 14;  
QY 31 LPEICT-----QCPGSVQNLKVAFYCK-----TTRE-----LMLHA---RCC-- 65  
DB 413 IAEFTGDSGSCGRKSGWVQCSKQDVLGCF-----LLCVNISGAPRLGDLVGISSVTF 525  
QY 66 -LNOKGT-----ILGLDQNLCSLEDPGPNFHQAHTTVIIDLQANPLKGLDLANTFRGFTQ 118  
DB 471 KLVNVEGTERGSCGRKSGWVQCSKQDVLGCF-----LLCVNISGAPRLGDLVGISSVTF 525

```

; TELEPHONE: (616) 381-115
;
; TELEFAX: (616) 381-5465
; INFORMATION FOR SEQ ID NO: 4
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
;

```

TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 LIBRARY: human fetal brain cDNA library  
 US-08-243-542-4

Query Match 8.0%; Score 99; DB 1; Length 769;  
 Best Local Similarity 21.4%; Pred. No. 0.076;  
 Matches 55; Conservative 32; Mismatches 90; Indels 80; Gaps 14;

OY 31 LPEICT-----OCPSVONLSKVAFYCK-----TTRE-----LMLHA---RCC-- 65  
 DB 512 IAEITGSSQCPNHLKLD--GYCDHEGRCYGRCKTRDRCQVLMGHAADRFCE 569  
 OY 66 -LNQKGT-----ILGLDLQNCSLDPGPNFQAHTTVIIDLANPLKGLDANFRGFTQ 118  
 DB 570 KLVNVEGTERGSGCRKSGWQCSKQDVLGCF-----LLCVNISGAPRLGDLVGDSSVTF 624  
 OY 119 LQTLILPQHVNCPG--INAMNTITSYID-----NOICGQKNL-----CNNTG 160  
 DB 625 YHQ---GKELDCRGHVQLADGSDLSYVEDGTACGPNMLCLDHRCPLPASAFNFSTCPGSG 681  
 OY 161 DPMCPENGSCVPDGPGLQCVADGDFHGYKC-----MRGGSFSLMFF 204  
 DB 682 ERRICSHHGVCNKG---KCICQPDWTKDCSINHPLTSPPTGTERYKGPSTNTIIT 737  
 OY 205 GILGATTLVSILLMAT 221  
 DB 738 GSAGAVLVAIVLGCT 754

RESULT 10  
 US-08-477-407-4  
 Sequence 4, Application US/08477407  
 Patent No. 5631351

GENERAL INFORMATION:  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: EMI, MITSURU  
 TITLE OF INVENTION: MDC PROTEINS AND DNAs  
 TITLE OF INVENTION: ENCODING THE SAME  
 NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FLINN, THEIL, BOUTELL & TANIS P.C.  
 STREET: 2026 Rambling Road

CITY: Kalamazoo  
 STATE: Michigan  
 COUNTRY: USA

ZIP: 49008-1699

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage

COMPUTER: IBM PC/XT/AT Compatible

OPERATING SYSTEM: MS-DOS 5.0

SOFTWARE: Wordperfect 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,407

FILING DATE: 07-JUN-1995

CLASSIFICATION: 530

Prior Application DATA:

APPLICATION NUMBER: US 08/243,542

FILING DATE: 13-MAY-1994

APPLICATION NUMBER: JP 5-136602

FILING DATE: 14 MAY 1993

APPLICATION NUMBER: JP 5-257455

FILING DATE: 22 SEPTEMBER 1993

APPLICATION NUMBER: JP 6-49904

FILING DATE: 23 FEBRUARY 1994

APPLICATION NUMBER: JP 6-73328

FILING DATE: 12 APRIL 1994

APPLICATION NUMBER: JP 6-84470

FILING DATE: 22 APRIL 1994

ATTORNEY/AGENT INFORMATION:  
 NAME: Terrylene F. Chapman  
 REGISTRATION NUMBER: 32 549  
 REFERENCE/DOCKET NUMBER: Furuya Case 1313  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (616) 381-1156  
 TELEFAX: (616) 381-5465  
 INFORMATION FOR SEQ. ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 769 amino acids  
 TYPE: amino acid  
 MOLECULE TYPE: linear  
 TOPOLOGY: linear  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 LIBRARY: human fetal brain cDNA library  
 US-08-477-407-4

Query Match 8.0%; Score 99; DB 1; Length 769;  
 Best Local Similarity 21.4%; Pred. No. 0.076;  
 Matches 55; Conservative 32; Mismatches 90; Indels 80; Gaps 14;

OY 31 LPEICT-----OCPSVONLSKVAFYCK-----TTRE-----LMLHA---RCC-- 65  
 DB 512 IAEITGSSQCPNHLKLD--GYCDHEGRCYGRCKTRDRCQVLMGHAADRFCE 569  
 OY 66 -LNQKGT-----ILGLDLQNCSLDPGPNFQAHTTVIIDLANPLKGLDANFRGFTQ 118  
 DB 570 KLVNVEGTERGSGCRKSGWQCSKQDVLGCF-----LLCVNISGAPRLGDLVGDSSVTF 624  
 OY 119 LQTLILPQHVNCPG--INAMNTITSYID-----NOICGQKNL-----CNNTG 160  
 DB 625 YHQ---GKELDCRGHVQLADGSDLSYVEDGTACGPNMLCLDHRCPLPASAFNFSTCPGSG 681  
 OY 161 DPMCPENGSCVPDGPGLQCVADGDFHGYKC-----MRGGSFSLMFF 204  
 DB 682 ERRICSHHGVCNKG---KCICQPDWTKDCSINHPLTSPPTGTERYKGPSTNTIIT 737  
 OY 205 GILGATTLVSILLMAT 221  
 DB 738 GSAGAVLVAIVLGCT 754

RESULT 11  
 US-08-484-355-4  
 Sequence 4, Application US/08484355  
 Patent No. 5705341

GENERAL INFORMATION:  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: EMI, MITSURU

TITLE OF INVENTION: MDC PROTEINS AND DNAs

TITLE OF INVENTION: ENCODING THE SAME

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLINN, THEIL, BOUTELL & TANIS P.C.

STREET: 2026 Rambling Road

CITY: Kalamazoo

STATE: Michigan

COUNTRY: USA

ZIP: 49008-1699

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage

COMPUTER: IBM PC/XT/AT Compatible

OPERATING SYSTEM: MS-DOS 5.0

SOFTWARE: Wordperfect 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,355

FILING DATE: 07-JUN-1995

CLASSIFICATION: 536

Prior Application DATA:

APPLICATION NUMBER: US 08/243,542

FILING DATE: 13-MAY-1994  
APPLICATION NUMBER: JP 5-136602  
FILING DATE: 14 MAY 1993  
APPLICATION NUMBER: JP 5-257455  
FILING DATE: 22 SEPTEMBER 1993  
APPLICATION NUMBER: JP 6-49904  
FILING DATE: 23 FEBRUARY 1994  
APPLICATION NUMBER: JP 6-73328  
FILING DATE: 12 APRIL 1994  
APPLICATION NUMBER: JP 6-84470  
FILING DATE: 22 APRIL 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Terryence F. Chapman  
REGISTRATION NUMBER: 32 549  
REFERENCE/DOCKET NUMBER: Futuya Case 1313  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (616) 381-1156  
TELEFAX: (616) 381-5465  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 769 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
LIBRARY: human fetal brain cDNA library  
US-08-484-355-4

Query Match 8.0%; Score 99; DB 1; Length 769;  
Best Local Similarity 21.4%; Pred. No. 0.076;  
Matches 55; Conservative 32; Mismatches 90; Indels 80; Gaps 14;

QY 31 LPEICT---QCPGSYONLSKAVTCK-----TTRE---LMLHA---RCC-- 65  
DB 512 IAEITGTGSSQCPNMLHKID--GYCDEHQRGRCYGRCKTRDROCOVLGMHAADRFCYE 569  
QY 66 -INOKGT-----ILGLDQNCLEDPGPNFHQAHTTYIIDQANPLKGLDLANTFRGTQ 118  
DB 570 KINVESTGSGCRKSGKSWQCSKODVLCG---LTCVNISGAPRLGDLVDISSVTF 624  
QY 119 IOTLLPQHVNCPG--INAMNTITSYID-----NOICOGOKNL-----CNNTG 160  
DB 625 YHQ---GKELDRCGHVYOLADGSDLSYVEDGTACGPNMLCDHRLPASAFTSCPSG 681  
QY 161 DEKMPKENGSCYPPDGPGLQCYCADGFHYKC-----MROGSFSLMFF 204  
DB 682 ERRICSHHGVCSENG---KCIQDPMTGKDCSINHPLPTSPPTGETERYKGPSTGNTIII 737  
QY 205 GILGATTLVSILLMT 221  
DB 738 GSIAQVLYAIVLGGT 754

RESULT 12  
PCT-US95-11684-2

Sequence 2, Application PC/TUS9511684  
GENERAL INFORMATION:  
APPLICANT: THE SCRIPPS RESEARCH INSTITUTE  
TITLE OF INVENTION: CYTOTOXIC DERIVATIVES THAT STIMULATE  
TITLE OF INVENTION: ATTACHMENT AND NEURITE OUTGROWTH, AND METHODS OF MAKING  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: The Scripps Research Institute, Office of  
ADDRESS: Patent Counsel  
STREET: 10666 North Torrey Pines Road, TPC 8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/11684  
FILING DATE: 14-SEP-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/308,359  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Logan, April C.  
REGISTRATION NUMBER: 33,950  
REFERENCE/DOCKET NUMBER: BEC0019P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-554-2937  
TELEFAX: 619-554-6312  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2199 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-11684-2

Query Match 7.8%; Score 96.5; DB 5; Length 2199;  
Best Local Similarity 22.8%; Pred. No. 0.58;  
Matches 55; Conservative 29; Mismatches 76; Indels 81; Gaps 17;

QY 7 GSLTLYPMAALLAAGVERAL-----ALPEICTQCPGSYONLSKAVT 51  
DB 2 GANTQLLAGVFLAFLALATEGGVLRKVIIRKROSGVNAFLPE-----EN-QPVFN 51  
QY 52 CKTREIMLARCCINOKGTILGLDQNCSE-----DPGNFHQAHTT-----VIID 99  
DB 52 HVTNRT DVSQOC-----SYDLESASGEKDLAPSPSPSF-QEHTVDEGNIVTF 101  
QY 100 LOAN-? C-----DLANTFRGTQTLILPQHVNCPGGIN-AMNTITSYIDNO-IC 149  
DB 102 HRINIPRRACGAAPADVKELLRLLEMLVSLRQCTAGAGCCIQPATGRDLTRPC 161  
QY 150 QGQKN-----LCN-----NIGDPKMPEN-----GSCVPDPGGLQCYCADGFHYGK 191  
DB 162 SGRGNFTSGCGVCPEGMKPMNCSEPE-CPGNCHLRGRCI-DG-----QCICDDGFTGED 215  
QY 192 C 19  
DB 216 C 21

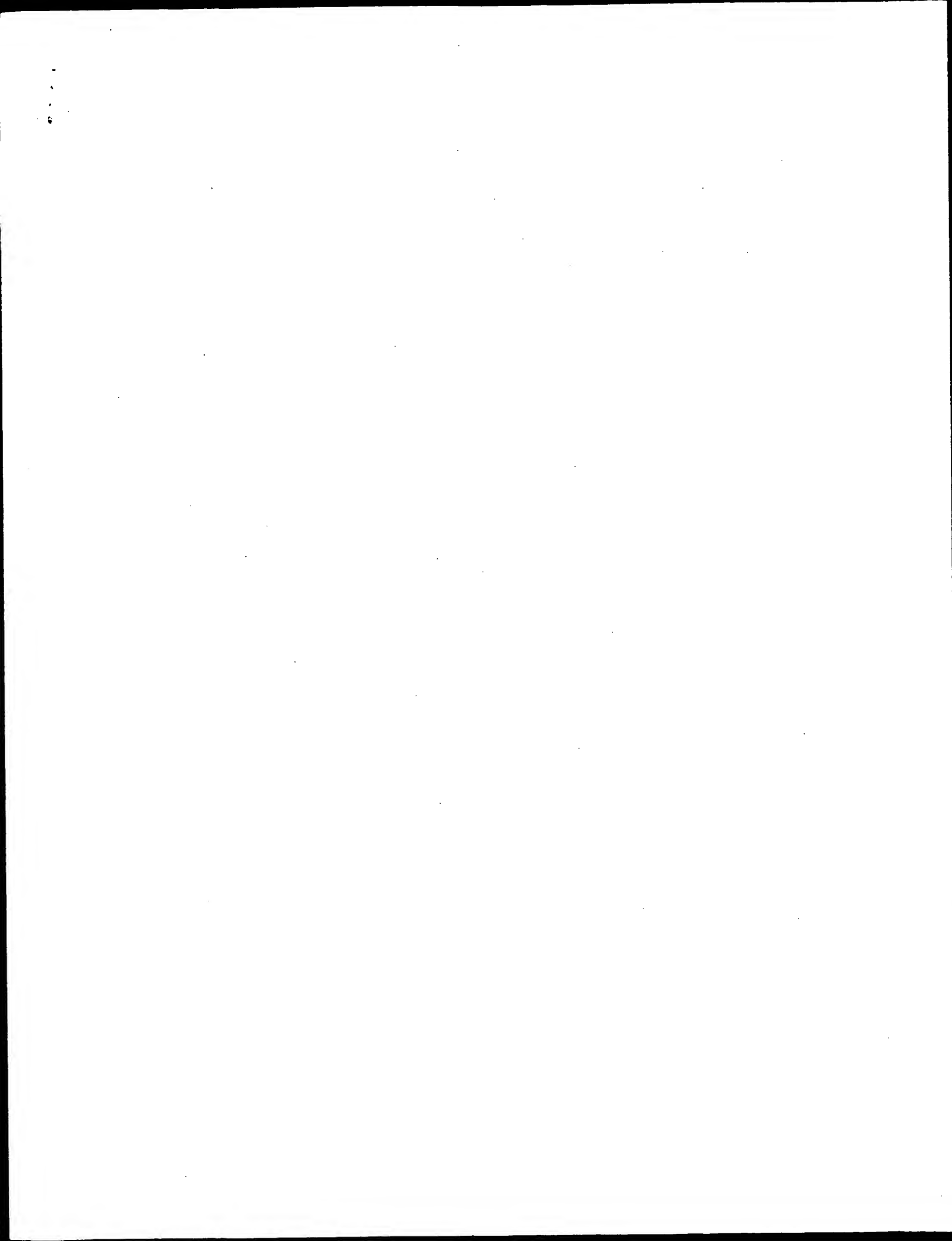
RESULT 13  
US-08-185-432-

Sequence 16, Application US/08185432  
Patent No. 5,796,612  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Busseau, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Watsuno, Kenji  
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

RESULT 14  
 US-08-083-590A-19  
 Sequence 19, Application US/08083590A  
 Patent No. 5786158  
 GENERAL INFORMATION:  
 APPLICANT: Artavanis-Tsakonas, S. et al.  
 TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
 TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/083,590A  
 FILING DATE: 25-JUN-1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 7320-015

RESULT 15  
 US-08-532-384-19  
 : Sequence 19, Application US/08532384  
 : Patent No. 6083904  
 : GENERAL INFORMATION:  
 : APPLICANT: Artavanis-Tsakonas, S. et al.  
 : TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
 : TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And  
 : TITLE OF INVENTION: Nucleic Acids  
 : NUMBER OF SEQUENCES: 21  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Pennie & Edmonds  
 : STREET: 1155 Avenue of the Americas  
 : CITY: New York  
 : STATE: New York  
 : COUNTRY: U.S.A.  
 : ZIP: 10036  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patentin Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/532,384  
 : FILING DATE:  
 : CLASSIFICATION: 424  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 08/083,590  
 : FILING DATE: 25-JUN-1993  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Mistrock, S. Leslie  
 : REGISTRATION NUMBER: 18,872  
 : REFERENCE/DOCKET NUMBER: 7326-015  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 212 790-9090  
 : TELEFAX: 212 869864/9741  
 : TELEX: 66141 PENNIE  
 : INFORMATION FOR SEQ ID NO: 19:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 2471 amino acids  
 : TYPE: amino acid  
 : STRADEDNESS: single







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:09:06 ; Search time 32.18 Seconds

(without alignments)  
160.138 Million cell updates/sec

Title: US-09-726-348-2

Perfect score: 229  
Sequence: 1 MAPHGPSTITLVPMMAALL.....TTLVSILLMATQRKAKTS 229

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size: 0

Total number of hits satisfying chosen parameters: 132412

Minimum DB seq length: 0  
Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database:

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	2.6	8	1 US-08-633-760-3	Sequence 3, Appl
2	6	2.6	18	2 US-08-764-640-42	Sequence 42, Appl
3	6	2.6	18	3 US-08-973-225-42	Sequence 42, Appl
4	6	2.6	18	3 US-09-244-298A-42	Sequence 42, Appl
5	6	2.6	18	4 US-09-516-704-42	Sequence 42, Appl
6	6	2.6	19	1 US-07-679-052A-1	Sequence 1, Appl
7	6	2.6	21	1 US-07-679-052A-9	Sequence 9, Appl
8	6	2.6	25	2 US-08-997-080-1	Sequence 1, Appl
9	6	2.6	25	3 US-08-997-362-1	Sequence 1, Appl
10	6	2.6	25	3 US-08-873-970-1	Sequence 1, Appl
11	6	2.6	25	4 US-09-095-855-1	Sequence 1, Appl
12	6	2.6	25	4 US-08-705-347A-1	Sequence 1, Appl
13	6	2.6	26	3 US-08-912-272-44	Sequence 44, Appl
14	6	2.6	27	4 US-08-968-542C-34	Sequence 34, Appl
15	6	2.6	29	4 US-09-042-353-295	Sequence 295, Appl
16	6	2.6	30	4 US-08-758-417A-143	Sequence 143, Appl
17	6	2.6	39	4 US-09-105-390-20	Sequence 20, Appl
18	5	2.2	6	1 US-07-718-577-17	Sequence 17, Appl
19	5	2.2	7	1 US-08-166-930-17	Sequence 17, Appl
20	5	2.2	7	2 US-08-727-045A-17	Sequence 17, Appl
21	5	2.2	8	4 US-07-861-458C-115	Sequence 115, Appl
22	5	2.2	9	1 US-07-908-317-1	Sequence 1, Appl
23	5	2.2	9	1 US-08-178-570-22	Sequence 22, Appl
24	5	2.2	9	1 US-08-615-181-94	Sequence 94, Appl
25	5	2.2	9	2 US-08-194-981E-6	Sequence 6, Appl
26	5	2.2	9	3 US-08-369-643-22	Sequence 22, Appl
27	5	2.2	9	3 US-08-159-339A-541	Sequence 541, Appl

28	5	2.2	9	3 US-08-159-339A-697	Sequence 697, Appl
29	5	2.2	9	5 PCT-US93-06171-1	Sequence 1, Appl
30	5	2.2	9	5 PCT-US95-00147-22	Sequence 22, Appl
31	5	2.2	10	1 US-08-250-789A-69	Sequence 69, Appl
32	5	2.2	10	1 US-08-250-789A-92	Sequence 92, Appl
33	5	2.2	10	1 US-08-346-333-46	Sequence 46, Appl
34	5	2.2	10	1 US-08-503-062-18	Sequence 18, Appl
35	5	2.2	10	2 US-08-724-548-24	Sequence 24, Appl
36	5	2.2	10	2 US-08-724-548-25	Sequence 25, Appl
37	5	2.2	10	2 US-08-724-548-26	Sequence 26, Appl
38	5	2.2	10	2 US-08-724-548-27	Sequence 27, Appl
39	5	2.2	10	2 US-08-724-548-28	Sequence 28, Appl
40	5	2.2	10	2 US-08-724-548-29	Sequence 29, Appl
41	5	2.2	10	2 US-08-764-640-59	Sequence 59, Appl
42	5	2.2	10	2 US-08-194-981E-9	Sequence 9, Appl
43	5	2.2	10	3 US-07-978-674B-24	Sequence 24, Appl
44	5	2.2	10	3 US-07-978-674B-25	Sequence 25, Appl
45	5	2.2	10	3 US-07-978-674B-26	Sequence 26, Appl

#### ALIGNMENTS

RESULT 1  
US-08-633-760-3  
; Sequence 3, Application US/08633760  
; Patent No. 5804429  
; GENERAL INFORMATION:  
; APPLICANT: NITTA, MINO  
; APPLICANT: SATO, YOSHIMASA  
; APPLICANT: FUJIMURA, TAKAO  
; APPLICANT: ISHII, YOSHIMORI  
; APPLICANT: NOGUCHI, YUJI  
; TITLE OR INVENTION: A NEW CEPHALOSPORIN C ACYLASE  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/633,760  
; FILING DATE: 01-MAY-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 18-929-0 PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 3:  
; LENGTH: 8 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-633-760-3

Query Match 2.6%; Score 6; DB 1; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 27 RALALP 32  
 |||||  
 Db 3 RALALP 8

# RESULT 2

US-08-764-640-42  
 ; Sequence 42, Application US/08764640  
 ; Patent No. 5869451  
 ; Patent No. 5869451 5837683

## GENERAL INFORMATION:

APPLICANT: Dower, William J.  
 APPLICANT: Barrett, Ronald W.  
 APPLICANT: Cwirla, Steven E.  
 APPLICANT: Gates, Christian  
 APPLICANT: Schatz, Peter J.  
 APPLICANT: Balasubramanian, Palaniappan  
 APPLICANT: Wagstrom, Christopher R.  
 APPLICANT: Hendren, Richard W.  
 APPLICANT: Depurtine, Randolph B.  
 APPLICANT: Podduturi, Surekha  
 APPLICANT: Yin, Qun  
 TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A  
 TITLE OF INVENTION: RECEPTOR  
 NUMBER OF SEQUENCES: 244  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Glaxo Wellcome  
 STREET: Five Moore Drive, P.O. Box 13398  
 CITY: Research Triangle Park  
 STATE: NC  
 COUNTRY: USA  
 ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/764,640  
 FILING DATE: 11-DEC-1996  
 CLASSIFICATION: 514

## ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.  
 REGISTRATION NUMBER: 36,392  
 REFERENCE/DOCKET NUMBER: PK3281  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-248-1000  
 INFORMATION FOR SEQ ID NO: 42:

## SEQUENCE CHARACTERISTICS:

LENGTH: 18 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear

## MOLECULE TYPE: peptide

US-08-764-640-42

Query Match 2.6%; Score 6; DB 2; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 GSFSIL 201  
 |||||  
 Db 12 GSFSIL 17

# RESULT 3

US-08-973-225-42  
 ; Sequence 42, Application US/08973225A  
 ; Patent No. 6083913  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dower, William J.

Barrett, Ronald W.  
 Cwirla, Steven E.  
 Duffin, David J.  
 Gates, Christian  
 Haselden, Sherrill S.  
 Matheakis, Larry C.  
 Schatz, Peter J.  
 Wagstrom, Christopher R.  
 Wrighton, Nicholas C.

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A  
 TITLE OF INVENTION: THROMBOPOIETIN RECEPTOR

## NUMBER OF SEQUENCES: 232

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome  
 STREET: Five Moore Drive, P.O. Box 13398  
 CITY: Research Triangle Park  
 STATE: NC  
 COUNTRY: USA  
 ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/973,225A  
 FILING DATE: 04-DEC-1997

## ATTORNEY/AGENT INFORMATION:

NAME: Hrubiec, Robert T.  
 REGISTRATION NUMBER: 36,392  
 REFERENCE/DOCKET NUMBER: PK3065USW  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 919-248-1000  
 INFORMATION FOR SEQ ID NO: 42:

## SEQUENCE CHARACTERISTICS:

LENGTH: 18 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: <unknown>  
 TOPOLOGY: linear

## MOLECULE TYPE: peptide

US-08-973-225-42

Query Match 2.6%; Score 6; DB 3; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 196 GSFSIL 201  
 |||||  
 Db 12 GSFSIL 17

# RESULT 4

US-09-244-298A-42  
 ; Sequence 42, Application US/09244298A  
 ; Patent No. 6121238

## GENERAL INFORMATION:

APPLICANT: Dower, William J.  
 APPLICANT: Barrett, Ronald W.  
 APPLICANT: Cwirla, Steven E.  
 APPLICANT: Gates, Christian  
 APPLICANT: Schatz, Peter J.  
 APPLICANT: Balasubramanian, Palaniappan  
 APPLICANT: Wagstrom, Christopher R.  
 APPLICANT: Hendren, Richard W.  
 APPLICANT: Depurtine, Randolph B.  
 APPLICANT: Podduturi, Surekha  
 APPLICANT: Yin, Qun

TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A  
 TITLE OF INVENTION: RECEPTOR

NUMBER OF SEQUENCES: 244  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Glaxo Wellcome  
STREET: Five Moore Drive, P.O. Box 13398  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/244,298A  
FILING DATE: 11-DEC-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Hrubiec, Robert T.  
REGISTRATION NUMBER: 36,392  
REFERENCE/DOCKET NUMBER: PK3281  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-248-1000  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-244-298A-42

Query Match 2.6%; Score 6; DB 3; Length 18;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 GSFSLL 201  
Db 12 GSFSLL 17

RESULT 5  
US-09-516-704-42  
Sequence 42, Application US/09516704  
Patent No. 6251864  
GENERAL INFORMATION:  
APPLICANT: Dower, William J.  
Barrett, Ronald W.  
Cwirla, Steven E.  
Gates, Christian  
Schatz, Peter J.  
Balasubramanian, Palaniappan  
Wagstrom, Christopher R.  
Hendren, Richard W.  
Deprince, Randolph B.  
Poddaturi, Surekha  
TITLE OF INVENTION: PEPTIDES AND COMPOUNDS THAT BIND TO A  
RECEPTOR  
NUMBER OF SEQUENCES: 244  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Glaxo Wellcome  
STREET: Five Moore Drive, P.O. Box 13398  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/516,704  
FILING DATE: 01-Mar-2000

CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Hrubiec, Robert T.  
REGISTRATION NUMBER: 36,392  
REFERENCE/DOCKET NUMBER: PK3281  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-248-1000  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 42:  
US-09-516-704-42

Query Match 2.6%; Score 6; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 GSFSLL 201  
Db 12 GSFSLL 17

RESULT 6  
US-07-679-052A-1  
Sequence 1, Application US/07679052A  
Patent No. 5298400  
GENERAL INFORMATION:  
APPLICANT: WHITEFIELD, Peter L.  
RICHARDSON, Michael A.  
APPLICANT: EUNN, Clive L.  
TITLE OF INVENTION: RECOMBINANT PRODUCT  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 2213-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/679,052A  
FILING DATE: 19910506  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: E. Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16786/147 CHAC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: 703/683-4109  
TELEX: 499149  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
FEATURES:  
NAME/KEY: Peptide  
LOCATION: 1..19  
OTHER INFORMATION: /note="Synthetic signal peptide"  
US-07-679-052A-1

Query Match 2.6%; Score 6; DB 1; Length 19;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LILAIG 24  
|||||  
DB 4 LILAIG 9

## RESULT 7

US-07-679-052A-9  
Sequence 9, Application US/07679052A  
Patent No. 5298400  
GENERAL INFORMATION:  
APPLICANT: WHITEFIELD, Peter L.  
APPLICANT: RICHARDSON, Michael A.  
APPLICANT: BUNN, Clive L.  
TITLE OF INVENTION: RECOMBINANT PRODUCT  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/679,052A  
FILING DATE: 19910506  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16786/147 CHAC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ. ID NO.: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-679-052A-9

Query Match 2.6%; Score 6; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LILAIG 24  
|||||  
DB 4 LILAIG 9

## RESULT 8

US-08-997-080-1  
Sequence 1, Application US/08997080  
Patent No. 5968524  
GENERAL INFORMATION:  
APPLICANT: WATSON, JAMES D.  
APPLICANT: TAN, PAUL L.J.  
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,080  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Steath, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000.1007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:

INFORMATION FOR SEQ. ID NO.: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 25 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-997-080-1

Query Match 2.6%; Score 6; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 VPDGP 177  
|||||  
DB 14 VPDGP 19

## RESULT 9

US-08-997-362-1  
Sequence 1, Application US/0897362  
Patent No. 5985287  
GENERAL INFORMATION:  
APPLICANT: Tan, Paul  
APPLICANT: Miyama, Jun  
APPLICANT: Visser, Elizabeth  
APPLICANT: Skinner, Margot  
APPLICANT: Scott, Linda  
APPLICANT: Prestidge, Ross  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR  
NUMBER OF SEQUENCES: 194  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/997,362  
FILING DATE:

```

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-997-362-1

Query Match      2.6%  Score 6;  DB 2;  Length 25;
Best Local Similarity 100.0%;  Pred. No. 38;
Matches 6;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy  172 VPDGPG 177
    |||||
Db  14 VPDGPG 19

RESULT 10
US-08-873-970-1
Sequence 1, Application US/08873970
Patent No. 6001361
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyma, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
NUMBER OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,970
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

```

```

TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-873-970-1

Query Match      2.6%  Score 6;  DB 3;  Length 25;
Best Local Similarity 100.0%;  Pred. No. 38;
Matches 6;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy  172 VPDGPG 177
    |||||
Db  14 VPDGPG 19

RESULT 11
US-09-095-855-1
Sequence 1, Application US/09095855
Patent No. 6160093
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
NUMBER OF INVENTION: Treatment and diagnosis of Mycobacterial Infections
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-095-855-1

```

Query Match 2.6%; Score 6; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 172 VPDGPG 177  
|||||  
Db 14 VPDGPG 19

RESULT 12  
US-08-705-347A-1  
; Sequence 1, Application US/08705347A  
; Patent No. 6284255  
; GENERAL INFORMATION:  
; APPLICANT: Tan, Paul  
; APPLICANT: Miyama, Jun  
; APPLICANT: Visser, Elizabeth  
; APPLICANT: Skinner, Margot  
; APPLICANT: Scott, Linda  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF MYCOBACTERIAL INFECTIONS  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Speckman Picard PLLC  
; STREET: 2601 Elliott Avenue, Suite 4185  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/705,347A  
; FILING DATE: 28-AUG-1996  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sleath, Janet  
; REGISTRATION NUMBER: 37,007  
; REFERENCE/DOCKET NUMBER: 11000.1002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206.269.0565  
; TELEFAX: 206.269.0563  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-705-347A-1

Query Match 2.6%; Score 6; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 172 VPDGPG 177  
|||||  
Db 14 VPDGPG 19

RESULT 13  
US-08-912-272-44  
; Sequence 44, Application US/08912272  
; Patent No. 6093874  
; GENERAL INFORMATION:  
; APPLICANT: Jotoku, K. Diane  
; APPLICANT: Okamoto, Jack K.  
; TITLE OF INVENTION: Methods for Improving Seeds  
; NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/912,272  
; FILING DATE: 15-AUG-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/879,827  
; FILING DATE: 20-JUN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/700,152  
; FILING DATE: 20-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-067220US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Peptide  
; LOCATION: 1..26  
; OTHER INFORMATION: /note="RAP2.7 linker region"  
; US-08-912-272-44

Query Match 2.6%; Score 6; DB 3; Length 26;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 42 YONLSK 47  
|||||  
Db 4 YONLSK 9

RESULT 14  
US-08-968-542C-34  
; Sequence 34, Application US/08968542C  
; Patent No. 5981728  
; GENERAL INFORMATION:  
; APPLICANT: Myers, et al.  
; TITLE OF INVENTION: dulla Codes For A No. 5981728el Starch  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McGregor & Adler, LLP  
; STREET: 8011 Candle Lane  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh  
; SOFTWARE: Microsoft Word 6.0.1 for Macintosh

;;  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/968,542C  
; FILING DATE: No. 5981728ember 12, 1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benjamin Aaron Adler, Ph.D., J.D.  
; REGISTRATION NUMBER: 35,423  
; REFERENCE/DOCKET NUMBER: D6036  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (713) 777-2321  
; TELEFAX: (713) 777-6908  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 27 amino acid residues  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
; DESCRIPTION: amino acid  
; HYPOTHETICAL: no  
; ANTI-SENSE: no  
; FRAGMENT TYPE: internal  
; US-08-968-542C-34

Query Match 2.6%; Score 6; DB 2; Length 27;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 ILGLDL 77  
| | | | |  
Db 17 ILGLDL 22

RESULT 15  
US-09-042-353-295  
; Sequence 295, Application US/09042353  
; Patent No. 6255458  
; GENERAL INFORMATION:  
; APPLICANT: Lonberg, Nils  
; APPLICANT: Kay, Robert M.  
; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for  
; TITLE OF INVENTION: Producing Heterologous Antibodies  
; NUMBER OF SEQUENCES: 421  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/042,353  
; FILING DATE: 13-MAR-1998  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/810,279  
; FILING DATE: 17-DEC-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/853,408  
; FILING DATE: 18-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/904,068  
; FILING DATE: 23-JUN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/990,860

;;  
; FILING DATE: 16-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/053,131  
; FILING DATE: 26-APR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/096,762  
; FILING DATE: 22-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/155,301  
; FILING DATE: 18-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,739  
; FILING DATE: 03-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/165,699  
; FILING DATE: 10-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/209,741  
; FILING DATE: 09-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/352,322  
; FILING DATE: 07-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/544,404  
; FILING DATE: 10-OCT-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/728,463  
; FILING DATE: 10-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US96/16433  
; FILING DATE: 10-OCT-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/758,417  
; FILING DATE: 02-DEC-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US97/21803  
; FILING DATE: 01-DEC-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Apple, Randolph T.  
; REGISTRATION NUMBER: 36,429  
; REFERENCE/DOCKET NUMBER: 014643-0090400S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 295:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; STRAND: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-09-042-353-295

Query Match 2.6%; Score 6; DB 4; Length 29;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 16 AVALLL 21  
| | | | |  
Db 7 AVALLL 12

Search completed: March 28, 2002, 09:16:38  
Job time: 452 sec

Thu Mar 28 09:21:30 2002

us-09-726-348-2.rai

Page 8



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:13:31 ; Search time 135.07 Seconds  
(with about 31 arguments)

```
Title: US-09-726-348-2_COPY_1_177
Perfect score: 971
Sequence: 1 MAPHGPSLTTLVMAALL.....NTGPEMCPENGSCVDPDG 177
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%

Database : SPTREMBL\_17:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Length	DB	ID	Description
	1	716	73.7	208	4	Q9Y5L7	Q9Y5I7 homo sapien
	2	673.5	69.4	223	11	Q9DD14	Q9dd14 mus musculus
	3	668	68.8	171	4	Q9Y2R7	Q9y2r7 homo sapien
	4	106.5	11.0	2906	11	Q9WU09	Q9wu09 ratus norvegicus
	5	99	10.2	1064	10	Q9FX19	Q9fx19 arbidopsis
	6	92.5	9.5	3857	11	Q98840	Q98840 mus musculus
	7	91	9.4	2872	11	Q9WU08	Q9wu08 ratus norvegicus
	8	89.5	9.2	1012	10	Q9L8Z5	Q9lkz5 glycine max
	9	87	9.0	784	10	Q9L8Z9	Q9l8z9 arbidopsis
	10	86.5	8.9	255	11	Q9ZOM1	Q9zom1 mus musculus
	11	86.5	8.9	255	11	Q9JIS2	Q9jis2 mus musculus
	12	86.5	8.9	1134	10	Q65510	Q65510 arbidopsis
	13	85.5	8.8	389	5	Q21081	Q21081 cseonhabdi
	14	85.5	8.8	1012	10	Q9LKZ4	Q9lkz4 glycine max
	15	84	8.7	1008	10	Q9L8Z6	Q9lkz6 glycine max
	16	83	8.5	640	10	Q9CY98	Q9cy98 arbidopsis
	17	82.5	8.5	851	10	Q9SI06	Q9si06 arbidopsis
	18	82	8.4	970	10	Q9M6D8	Q9m6d8 cryza sativa
	19	82	8.4	992	10	Q65440	Q65440 arbidopsis

20	81.5	8.4	699	10	081765	081765 aradidopsis
21	81.5	8.4	1011	10	09S80	09S80 aradidopsis
22	81	8.3	943	10	09F643	09F643 aradidopsis
23	81	8.3	1079	10	09CA77	09CA77 aradidopsis
24	81	8.3	1515	13	09DE37	09DE37 brachydanio
25	80.5	8.3	1120	10	09LR24	09LR24 aradidopsis
26	80	8.2	601	4	094951	094951 homo sapien
27	80	8.2	708	13	P87363	P87363 gallus galli
28	80	8.2	879	10	081066	081066 aradidopsis
29	80	8.2	2062	4	09H231	09H231 homo sapien
30	79.5	8.2	358	11	P97946	P97946 mus musculus
31	79	8.1	1599	5	009983	009983 caenorhabdit
32	78.5	8.1	541	5	09W330	09W330 drosophila
33	78.5	8.1	787	10	09MTW1	09MTW1 aradidopsis
34	78.5	8.1	806	10	050025	050025 lycopersicon
35	78.5	8.1	1003	10	049545	049545 aradidopsis
36	78.5	8.1	1088	10	09LR66	09LR66 aradidopsis
37	78.5	8.1	1450	5	026627	026627 strongyloce
38	78	8.0	835	11	09OX05	09OX05 rattus norvig
39	78	8.0	979	10	09LPE0	09LPE0 aradidopsis
40	78	8.0	2931	5	09W2C6	09W2C6 drosophila
41	78	8.0	4599	11	09J118	09J118 mus musculus
42	77.5	8.0	643	5	09U051	09U051 nephrops
43	77	7.9	242	5	096687	096687 lumbricus b
44	77	7.9	766	10	023161	023161 aradidopsis
45	77	7.9	988	6	097867	097867 sus scrofa

## ALIGNMENTS

[illegible]



Query Match	11.0%;	Score 106.5;	DB 11;	Length 2906;
Best Local Similarity	23.2%;	Pred. No. 0.021;		
Matches 46;	Conservative 18;	Mismatches 71;	Indels 63;	Gaps 11.
QY	1	MARPGSGLTTLVPMAAALLALGVERALALPEICTQ--CPGSVONLSKVAFCYKTRREL	58	
DB	2428	ICPHPGGYATD-----GRDIDECKVWPSLCTNLGIC--VNTMGSRFCCKGYTT	2474	
QY	59	MLHARCCLNOKGTLLIGLDLONCSLEDDPGP-NPHQAHNT-----VIIDQLANPLKG	107	
DB	2475	DISGTAC-----VLDDECS-OSPPKPCNCTKNTBESSYCCSPRGVLAEDGKTC-	2523	
QY	108	DL-----AMTFRGFTLOTLTLPLQHNCPGSIAMWMTTSYIDNOICQCK	153	
DB	2524	DLDECGTQHONQOFLCVMTLIGFT-----CKSPRGFQHH--TACIDNNEGGSPR	2571	
QY	154	NICNNTGDPKPCRENGSC	171	
DB	2572	SLCGAKGICONTPGSFSC	2589	
RESULT	5			
ID	Q9FX19	PRELIMINARY;	PRT; 1064 AA.	
AC	Q9FX19;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	PURATIVE PROTEIN KINASE.			
CS	F12612.7.			
GN	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsids.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Federespiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,			
RA	Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,			
RA	Buehler E., Chao Q., Chin C., Chou J., Choi E., Gonzalez A.,			
RA	Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,			
RA	Tenzin C., Liu A., Liu S., Mukharly N., Pham P., Sakano H., Shinn P.,			
RL	Lotium M., Vaysberg M., Yu G., Becker J., Theologis A., Davis R.W.;			
DR	Submitted (Sep-2000) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AC015446; AAG12526.1;			
DR	InterPro: IPR001511; Aminoctran.1.			
DR	InterPro: IPR000719; Euk_pkinase.			
DR	InterPro: IPR001611; LRR.			
DR	InterPro: IPR003592; LRR_out.			
DR	InterPro: IPR002290; Ser_thr_kin_actsite.			
DR	InterPro: IPR001245; Tyr_kin.			
DR	Pfam: PF00560; LRR; 16.			
DR	Pfam: PF00069; pkinase; 1.			
DR	PRINTS: PR00019; LEURICHRPT.			
DR	SMART: SM00370; LRR; 23.			
DR	SMART: SM00221; STYK; 1.			
DR	SMART: SM00220; STYK; 1.			
DR	SMART: SM00219; TYRK; 1.			
DR	PROSITE: PS00105; AA_TRANSER_CLASS.1; UNKNOWN.1.			
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.			
KW	ATP-binding; Kinase; Transferase.			
SEQ	SEQUENCE 1064 AA; 115915 MW; 2E8FC24DFFAF053D CRC64;			
QY	37	QCPGVONLSKVAFCYKTRRELMLHARCCLNOKGTLLIGLDLONCSLEDDPGP-NPHQAHNTV	96	
DB	563	QIPKSKIKNL-----OKTLLDLSYNSLSGEIPEQLGQVYTSII	600	
Query Match	10.2%;	Score 99;	DB 10;	Length 1064;
Best Local Similarity	26.4%;	Pred. No. 0.046;		
Matches 37;	Conservative 14;	Mismatches 51;	Indels 38;	Gaps 4.

QY	97	IIDIDANPLKGLANTFREFLOLILPOHYNCGGINAMNTITSYIDNOICQGRKNC	156
Db	601	NIDISYNTFTGNIPTETFSLOLTOLOSULSSN-SLHGDIKVLGSLTSLASINIS-----C	653
QY	157	NNTGDPEMCPENGSCVPDGP	176
Db	654	NNFSGP-----IPSTP	664
RESULT 6			
ID	088840	PRELIMINARY:	PRT: 3857 AA.
AC	088840:		
DT	01-NOV-1998	(TREMBLrel. 08, Created)	
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)	
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	
DE	MUTANT FIBRILLIN-1.		
GN	FBNI.		
OS	Mus musculus (Mouse):		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=B10.D2;		
RX	MEDLINE=98069008; PubMed=9405934;		
RA	Bona C.A., Murai C., Casares S., Kasturi K., Nishimura H., Honjo T.,		
RA	Matsuda F.;		
RT	"Structure of the mutant fibrillin-1 gene in the tight skin (TSK)		
RT	mouse.";		
RL	DNA Res. 4:267-271(1997).		
DR	EMBL: AF007248; AAC62317.1. -		
DR	HSSP: P35555. 1AB1		
DR	InterPro: IPR000152; Asx_hydroxyl.		
DR	InterPro: IPR002557; Chitin_binding.		
DR	InterPro: IPR000561; EGF-like.		
DR	InterPro: IPR001881; EGF_Ca.		
DR	InterPro: IPR002212; TB.		
DR	InterPro: IPR000822; Znf-C2H2.		
DR	Pfam: PF00008; EGF; 64.		
DR	Pfam: PF00683; TB; 12.		
DR	SMART: SM00494; ChbD2; 2.		
DR	SMART: SM00179; EGF_CA; 60.		
DR	SMART: SM00001; EGF_like; 4.		
DR	PROSITE: PS00010; ASX_HYDROXYL; 61.		
DR	PROSITE: PS00022; EGF_1; UNKNOWN_2.		
DR	PROSITE: PS01186; EGF_2; 50.		
DR	PROSITE: PS01187; EGF_CA; 61.		
DR	PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.		
KW	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.		
SEQ	SEQUENCE 3857 AA; 418303 MW; 5BC061B8C527E04C CRC64;		
Query Match 9.5%; Score 92.5; DB 11; Length 3857;			
Best Local Similarity 20.9%; Pred. No. 0.95;			
Matches 42: Conservative 21; Mismatches 69; Indels 69; Gaps 9.			
QY	1	MAPHREGSITTLVPNAALLALGVRLALPEICT--QC---PGSVNLSKVA-----	49
Db	3375	LCPHRGGEFTN-----GADIDECKVIDHYCRNGECVNDRSYHCICITGTYTPDT	3424
QY	50	-----FYCKTTEELMHARCCLNOKGTIIGLDIQC-SLEDGPNPF	90
Db	3425	GTACVDLNCNAPKPCNTIKNTE-----GSYQSCSPKGYIIQEDGRSKIDDECATKH	3480
QY	91	QAHTTVIIDLQANPKGLDIAFTFRGFTLOLTLIPQHYNCPGGINAMNTITSYIDNOICQ	150
Db	3481	NCQFLCV-----NFIQFT-----CKSPGFTQHH--TACIDNNECT	3515
QY	151	GCKNLGNNTGDEMCPEMGSC	171
Db	3516	SDINLGSGSKGVQNTPGSFTC	3536

```

RESULT 7
O9MUB8      PRELIMINARY;      PRT;      2872 AA.
ID O9MUB8
AC O9MUB8;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE FIBRILLIN-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99032689; PubMed=9815129;
RA Kanwar Y.S., Ota K., Yang Q., Kumar A., Wada J., Kashihara N.,
RA Peterson D.R.;
RT "Isolation of rat fibrillin-1 cDNA and its relevance in metaphoric
RT development."
RL Am. J. Physiol. 275:F710-F723(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Kanwar Y.S.;
RX Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF135059; AAD34438.1;
DR HSSP: P35555; IAPJ.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002212; TB.
DR InterPro: IPR000822; ZnF-C2H2.
DR Pfam: PF00008; EGF_46.
DR Pfam: PF00683; TB; 9.
DR SMART: SM00179; EGF_CA_41.
DR SMART: SM00010; Asx_hydroxyl; 5.
DR PROSITE: PS00010; Asx_hydroxyl; 42.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 38.
DR PROSITE: PS01187; EGF_CA_41.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 2872 AA; 312069 MW; 0C4F73B87A80280 CRC64;

Query Match          9.4%; Score 91; DB 11; Length 2872;
Best Local Similarity 25.4%; Pred. No. 1;
Matches 35; Conservative 13; Mismatches 54; Indels 36; Gaps 6;

QY 35 CTGCGSVQNLKVAFYCKTTRRLMLHARCCINQGTILGLDQNC-SLEDPGFNFQAH 93
DB 2449 CNQAP-----KPCNFICKNTE---GSYQSCPKGYILDGSGSCDDECAIKQHNQ 2498
QY 94 TTVIIDLQANPLKGLDANTFRGFTQLQLLPQHVNCPCGINANNTTSTYDNOICGQK 153
DB 2499 FLVC-----NTIGFT-----CKCPPEFTQHH--TACIDNNECTSEI 2533
QY 154 NLGNNTGDEMCPCNGSC 171
DB 2534 NLGSGKVCQNTPGSFTC 2551

```

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamamoto E., Knap H.T.;
RT "Soybean receptor-like protein kinase (GmRLK2).";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL: AF244889; AAF91323.1;
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR InterPro: IPR001245; Tyr_kin.
DR InterPro: IPR000130; Zn_MTpeptidse.
DR Pfam: PF00560; LRR; 14.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00019; LEURICHRPT.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00370; LRR; 17.
DR SMART: SM00221; STYK; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1012 AA; 110115 MW; F8C74BB99DB7E0E6 CRC64;

Query Match          9.2%; Score 89.5; DB 10; Length 1012;
Best Local Similarity 25.5%; Pred. No. 0.46;
Matches 39; Conservative 23; Mismatches 62; Indels 29; Gaps 9;

QY 41 SVONLSKVAFYCKTTRRLMLHARCCINQGTILG-----IDLQNCSTLEDP-GPNFQAH 92
DB 471 SIGNESSV-----OKLLDGMFTGRLEPTQIGRLQSLKIDPSGKRFSGPIAETISQC 523
QY 93 HTTVIIDLQANPLKGLDANTFRGFTQLQLLPQ-HV--NCPGINANNTTSTYDNO 147
DB 524 KLTFELDLRNELSGDIPREITGMRLYNLNSKHNLVGSISSMSQSTSVDSYNN 583
QY 148 IC-----QCGKNCNNT---GDEMC-PENGSC 171
DB 584 LSGLVGTGQFSYFNTSFLGNPDLCGYLAGC 616

RESULT 9
O9LS79      PRELIMINARY;      PRT;      784 AA.
ID O9LS79
AC O9LS79;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DISEASE RESISTANCE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:131-135(2000).

```

DR EMBL: AB026655; BAB02104.1; -  
 DR InterPro: IPR001611; LRR.  
 DR InterPro: IPR003592; LRR\_out.  
 DR Pfam: PF005560; LRR\_14.  
 DR PRINTS: PR00019; LEURICHRPT.  
 DR SMART: SM00370; LRR\_14.  
 SO SEQUENCE 784 AA; 87228 MW; 99922846B551A11 CRC64;

Query Match 9.0%; Score 87; DB 10; Length 784;  
 Best Local Similarity 20.5%; Pred. No. 0.65;  
 Matches 33; Conservative 26; Mismatches 60; Indels 42; Gaps 4;

QY 23 LGVERALALPEICTO-----CPGSVONLSKVAFYCKTTRRLMARCC 66  
 ID 0920M1 PRELIMINARY; PRT; 255 AA.  
 AC 0920M1;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
 DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.  
 GN K1K4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS-WEBSTER;  
 RA Sliemer J.;  
 RT "Enamel Matrix Serine Proteinase 1 (EMSP1).";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.  
 CC -1- SIMILARITY: TO CHYMOTRYPsin SERINE PROTEASE FAMILY (S1).  
 DR EMBL: AF019979; AAC98894.1; -  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.251; -  
 DR MGD: MGI:1861379; K1K4.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; TRYp-Spc.1.  
 DR PROSITE: PS00134; TRYpSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYpSIN\_SER; 1.  
 DR HydroLase: Serine protease; Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 32 255 MATRIX SERINE PROTEINASE 1.  
 SO SEQUENCE 255 AA; 27488 MW; 6BD2E7DEA060A2A CRC64;

RESULT 10  
 Q920M1 PRELIMINARY; PRT; 255 AA.  
 AC 0920M1;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
 DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.  
 GN K1K4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS-WEBSTER;  
 RA Sliemer J.;  
 RT "Enamel Matrix Serine Proteinase 1 (EMSP1).";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.  
 CC -1- SIMILARITY: TO CHYMOTRYPsin SERINE PROTEASE FAMILY (S1).  
 DR EMBL: AF019979; AAC98894.1; -  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.251; -  
 DR MGD: MGI:1861379; K1K4.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; TRYp-Spc.1.  
 DR PROSITE: PS00134; TRYpSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYpSIN\_SER; 1.  
 DR HydroLase: Serine protease; Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 32 255 MATRIX SERINE PROTEINASE 1.  
 SO SEQUENCE 255 AA; 27488 MW; 6BD2E7DEA060A2A CRC64;

Query Match 8.9%; Score 86.5; DB 11; Length 255;  
 Best Local Similarity 22.1%; Pred. No. 0.22;  
 Matches 51; Conservative 28; Mismatches 59; Indels 93; Gaps 14;

QY 2 APHGSGSLTLVPMALALLALGVERALALPEICTOCPGSVONLSKVAFYCK---TTRRL 58  
 ID 0920M1 PRELIMINARY; PRT; 255 AA.  
 AC 0920M1;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
 DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.  
 GN K1K4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS-WEBSTER;  
 RA Sliemer J.;  
 RT "Enamel Matrix Serine Proteinase 1 (EMSP1).";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.  
 CC -1- SIMILARITY: TO CHYMOTRYPsin SERINE PROTEASE FAMILY (S1).  
 DR EMBL: AF019979; AAC98894.1; -  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.251; -  
 DR MGD: MGI:1861379; K1K4.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; TRYp-Spc.1.  
 DR PROSITE: PS00134; TRYpSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYpSIN\_SER; 1.  
 DR HydroLase: Serine protease; Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 32 255 MATRIX SERINE PROTEINASE 1.  
 SO SEQUENCE 255 AA; 27488 MW; 6BD2E7DEA060A2A CRC64;

DB 67 VLSAARCL-DESYVIGLHNLKSGQEPGRMLEAHLSTQHPNPNDFPANDLMILKINE 125  
 QY 103 NPLKGLDANTFR-----GFTQLOTLLPQHVNCPG-GINAMWTITSY 143  
 DB 126 SVIE---SNTIRSIPIVATQCPPTPGDCLVSGWGQKNGKLPsLLCCVNLVSASEFCRL 182  
 QY 144 ID-----NOICOG---OKMLCN-NTGDPMCPEN-----GSCVDPG 175  
 DB 183 YDPVYHLSMFCAGGQDQKDCNGDSGGRFVNCNRSIQGLVSMGQKCGCPG 233

RESULT 11  
 Q9JIS2 PRELIMINARY; PRT; 255 AA.  
 AC 09JIS2;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
 DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.  
 GN K1K4 OR PRSS17  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SVJ;  
 RA MEDLINE=20323211; PubMed=10863090;  
 RX Hu J.C.-C., Zhang C., Sun X., Yang Y., Cao X., Ryu O., Sliemer J.P.;  
 RT "Characterization of the mouse and human PRSS17 genes, their relationship to other serine proteases, and the expression of PRSS17 in developing mouse incisors.";  
 RL Gene 251:1-8(2000).  
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYpSIN FAMILY.  
 CC -1- SIMILARITY: TO CHYMOTRYPsin SERINE PROTEASE FAMILY (S1).  
 DR EMBL: AF198031; AAF85937.1; -  
 DR MGD: MGI:1861379; K1K4.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; TRYp-Spc.1.  
 DR PROSITE: PS00134; TRYpSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYpSIN\_SER; 1.  
 DR HydroLase: Serine protease; Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 32 255 ENAMEL MATRIX SERINE PROTEINASE 1.  
 FT VARIANT 15 15 Y -> C.  
 FT VARIANT 251 251 T -> I.  
 SO SEQUENCE 255 AA; 27536 MW; CBF5FBAF7D4D679C CRC64;

Query Match 8.9%; Score 86.5; DB 11; Length 255;  
 Best Local Similarity 22.1%; Pred. No. 0.22;  
 Matches 51; Conservative 28; Mismatches 59; Indels 93; Gaps 14;

QY 2 APHGSGSLTLVPMALALLALGVERALALPEICTOCPGSVONLSKVAFYCK---TTRRL 58  
 ID 0920M1 PRELIMINARY; PRT; 255 AA.  
 AC 0920M1;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)  
 DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.  
 GN K1K4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS-WEBSTER;  
 RA Sliemer J.;  
 RT "Enamel Matrix Serine Proteinase 1 (EMSP1).";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.  
 CC -1- SIMILARITY: TO CHYMOTRYPsin SERINE PROTEASE FAMILY (S1).  
 DR EMBL: AF019979; AAC98894.1; -  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.251; -  
 DR MGD: MGI:1861379; K1K4.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; TRYp-Spc.1.  
 DR PROSITE: PS00134; TRYpSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYpSIN\_SER; 1.  
 DR HydroLase: Serine protease; Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 32 255 ENAMEL MATRIX SERINE PROTEINASE 1.  
 FT VARIANT 15 15 Y -> C.  
 FT VARIANT 251 251 T -> I.  
 SO SEQUENCE 255 AA; 27536 MW; CBF5FBAF7D4D679C CRC64;

RESULT 12  
ID 065510 PRELIMINARY: PRT: 1134 AA.  
AC 065510:  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE PUTATIVE RECEPTOR PROTEIN KINASE.  
GN F23E13.70 OR AT4G36180.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Beyer M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Hohnselt J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.,  
RA Schueller C.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,  
RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1 SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AL022141; CAI8124.1; -  
DR EMBL; AL161588; CAB81527.1; -  
DR HSSP; P00523; 2PRT.  
DR InterPro; IPR000719; Euk\_kinase.  
DR InterPro; IPR001611; LRR.  
DR InterPro; IPR003592; LRR\_out.  
DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
DR Pfam; PF00560; LRR; 21.  
DR PRINTS; PR00019; LEURICHRPT.  
DR SMART; SM00370; LRR; 25.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;  
KW transferase.  
SQ SEQUENCE 1134 AA; 123461 MW; 4480429380993AF8 CRC64;

Query Match 8.9%; Score 86.5; DB 10; Length 1134;  
Best Local Similarity 24.6%; Pred. No. 1.1;  
Matches 50; Conservative 27; Mismatches 91; Indels 35; Gaps 8;

OY 1 MAPHGGSLTTL-----VPMAAALLALGVERALALPEICTOCGSGSVQNLK 47  
DB 537 VPEGFSSLSRLRYVNLSSNFSFGELPQTFGLRL-VLSLSNHLSSGSIPEIGNSA 595  
OY 48 VAFYCKTRELMLARCLNKGTLIGDL--QNCSEDPENPHQATHTYIIDLQANPL 105  
DB 566 LEVLELRNRLMGLTPADLSRLPKLVLDLQNNLSGEIP-PEISQSSLSLSLDNHL 654  
OY 106 KGDLANPFGFTQLTLLIPQHVNCPGINA-----WNTITSYIDNQICQCKN 154  
DB 655 SGVIFGSGSLNLTLMKMLSVN-NLTGEIRASLALISSNLYFVSSNNLKGEIPASIGS 713  
OY 155 LCNMT---GDPEWC--PENGSC 171  
DB 714 RINTSEFGNTELCGKPLNRR 736

RESULT 13  
ID 021081 PRELIMINARY: PRT: 389 AA.  
AC 021081:  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE KOIC8.2 PROTEIN.  
GN KOIC8.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sims M.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Cosey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latelle P.,  
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rinken L., Koopra A., Saunders D., Showkeen R.,  
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
DR EMBL; Z49068; CAA8855.1; -  
DR InterPro; IPR002899; EB.  
DR SMART; SM00289; WRI; 7.  
SQ SEQUENCE 389 AA; 40335 MW; C32B38446991D17D CRC64;

Query Match 8.8%; Score 85.5; DB 5; Length 389;  
Best Local Similarity 25.3%; Pred. No. 0.44;  
Matches 41; Conservative 25; Mismatches 65; Indels 31; Gaps 11;

OY 21 LALGVERALALPEICTOCGSGSVONLSKVAFYCKTTEMLHARCLNKGTLIGDLQNC 80  
DB 188 IALITNGAL---EMCT-TPGT--QCSSAGYTCQLSVLATYVCCGQSGYSGTIG-----C 236  
OY 81 SLDDPGNFH---QAHTYIID-LQANPLKGLATFEGFQQLTL---LPQHVNC 131  
DB 237 A--DGRPVQIAGQITTEITSATISCPSSGYDCAPSDPEVDCCLGSTRIPENISCP 294  
OY 132 GGINAMNTITSYIDNQICOGKNLCNNTGDPENCGSCVP 173  
DB 295 TG---WNSTKNEVDNAV-----RTGFAVLDTG-CPIGYSCAP 327

RESULT 14  
ID 09LKZ4 PRELIMINARY: PRT: 1012 AA.  
AC 09LKZ4:  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE RECEPTOR-LIKE PROTEIN KINASE 3.  
GN RLK3.  
OS glycine max (soybean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamamoto E., Knapp H.T.;  
RT "Soybean receptor-like protein kinase (GmRLK3).";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1 SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.

```

RESULT 15
09LKZ6
ID 09LKZ6 PRELIMINARY; PRT; 1008 AA.
AC 09LKZ6
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE RECEPTOR-LIKE PROTEIN KINASE 1.
GN RLK1.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Oxycarpaceae; Fabiales; Fabaceae; Papilionoideae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamamoto E., Knap H.T.;
RT "Soybean receptor-like protein kinase (GmRLK1).";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDI databases.
CC 1-1. SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL, AF244888; AAF91322.1;
DR InterPro: IPR001511; Amlnothr_n_1.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR InterPro: IPR001245; Tyr_kin.
DR InterPro: IPR000130; Zn_MMPetase.
DR Pfam: PF00560; LRR_13.
DR Pfam: PF00069; pkinase_1.
DR PRINTS: PR00019; LEURICHAPT.
DR PRINTS: PR00109; TYRKINASE.

```

Search completed: March 28, 2002, 09:13:35  
Job time: 889 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:25:03 ; Search time 56.59 Seconds

(without alignments)  
457.505 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_1\_177  
Sequence: 1 MAPHGCSLITIVMAAAIL.....NTGDPMPCPDNGSCVPDGG 177

Scoring table: OLIGO

Searched: 473505 seqs, 146272329 residues

Word size: 0

Total number of hits satisfying chosen parameters: 12410

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Lasting first 45 summaries

Database:

SPTREMBL\_17:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mmc:  
8: sp\_organelle:  
9: sp-phage:  
10: sp-plant:  
11: sp-rodent:  
12: sp-virus:  
13: sp-vertebrate:  
14: sp\_unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	6	3.4	22	11	Q9WUK1	Q9WUK1 rattus norv
2	5	2.8	16	12	Q78377	Q78377 human immun
3	5	2.8	18	11	Q90MG4	Q90MG4 mus musculu
4	5	2.8	19	2	Q9R4A3	Q9R4A3 escherichia
5	5	2.8	20	6	Q9RVR8	Q9RVR8 oryctolagus
6	5	2.8	20	12	Q69381	Q69381 herpes simp
7	5	2.8	22	6	Q9TRC4	Q9TRC4 canis famill
8	5	2.8	22	12	Q68988	Q68988 human herpe
9	5	2.8	22	12	Q87084	Q87084 pseudorabie
10	5	2.8	23	2	Q9R5K9	Q9R5K9 bordetella
11	5	2.8	23	2	Q62594	Q62594 rattus leuc
12	5	2.8	23	11	Q62594	Q62594 rattus leuc
13	5	2.8	24	8	Q9TJ21	Q9TJ21 nicotiana t
14	5	2.8	25	4	Q9UL44	Q9UL44 homo sapien
15	5	2.8	25	12	Q99BS2	Q99BS2 human immun
16	5	2.8	25	12	Q99BR8	Q99BR8 human immun
17	5	2.8	26	4	Q9Z424	Q9Z424 pseudomonas
18	5	2.8	26	4	Q9NP34	Q9NP34 homo sapien
19	5	2.8	26	13	Q9PS44	Q9PS44 gallus gall

20	5	2.8	26	13	Q9PS43	Q9PS43 gallus gall
21	5	2.8	27	2	Q9S6D8	Q9S6D8 escherichia
22	5	2.8	27	6	Q29019	Q29019 sus scrofa
23	5	2.8	27	12	Q9Y205	Q9Y205 hepatitis c
24	5	2.8	27	13	Q9PRY4	Q9PRY4 gallus gall
25	5	2.8	28	8	Q9YIS4	Q9YIS4 leishmania
26	5	2.8	28	11	Q62677	Q62677 rattus norv
27	5	2.8	30	10	Q44171	Q44171 anabena sp
28	5	2.8	30	10	Q9S801	Q9S801 chlamydomon
29	4	2.3	8	11	Q9ET18	Q9ET18 mus spretus
30	4	2.3	8	11	Q9ET17	Q9ET17 mus caroll
31	4	2.3	8	11	Q9ET16	Q9ET16 mesocricetu
32	4	2.3	9	7	Q9MWA3	Q9MWA3 homo sapien
33	4	2.3	9	13	Q9TKD9	Q9TKD9 pericalymma
34	4	2.3	9	13	P82075	P82075 littoria rub
35	4	2.3	9	13	P82093	P82093 littoria rub
36	4	2.3	10	4	Q9H3R9	Q9H3R9 homo sapien
37	4	2.3	10	9	Q9TOR6	Q9TOR6 sus scrofa
38	4	2.3	11	6	Q9TRX0	Q9TRX0 bacterioph
39	4	2.3	11	7	Q9UEX7	Q9UEX7 polyomaviru
40	4	2.3	11	12	Q84247	Q84247 littoria rub
41	4	2.3	11	13	P82074	P82074 streptococ
42	4	2.3	12	2	Q9S550	Q9S550 epheida sp.
43	4	2.3	12	8	P92457	P92457 mycobacteri
44	4	2.3	13	2	Q50038	Q50038 mycobacteri
45	4	2.3	13	7	Q19789	Q19789 homo sapien

## ALIGNMENTS

RESULT	ID	Q9WUK1	PRELIMINARY:	PRT:	22 AA.
AC	Q9WUK1				
DT	01-NOV-1999	(TREMBLrel. 12, Created)			
DT	01-NOV-1999	(TREMBLrel. 12, last sequence update)			
DT	01-NOV-1999	(TREMBLrel. 12, last annotation update)			
DE	RELAXIN-LIKE FACTOR (FRAGMENT).				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_Taxid=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SPRAGUE-DAWLEY;				
RA	Spleen A.-N., Pusch W., Iwell R.;				
RT	"Cloning and sequencing of the rat relaxin-like factor and its				
RT	promotor";				
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF139920; A033851.1;				
FT	NON_TER	22			
FT	SEQUENCE	22 AA; 2325 MW; FLAAE6B6118F143F CRC64;			

Query Match	3.48;	Score 6;	DB 11;	Length 22;
Best Local Similarity	100.0%;	Pred. No. 99;		
Matches	6;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
			0;	
OY	19	LLIAG	24	
Db	8	LLIAG	13	
RESULT	2			
Q78377				
AC	Q78377	PRELIMINARY:	PRT:	16 AA.
ID	Q78377;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, last sequence update)		
DT	01-NOV-1998	(TREMBLrel. 08, last annotation update)		
DE	VIRAL SAMPLE FLIPBAC (FLORIDA PATIENT B), PARTIAL ENV CDS, V4 REGION			
	(FRAGMENT).			

OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zhang L.Q., Leigh-Brown A.J.;  
 RL Submitted (Apr-1992) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92271245; PubMed=1589796;  
 RA Ou C.Y., Ciesielski C.A., Myers G., Banea C.I., Luo C.C.,  
 RA Korber B.T.M., Mullins J.I., Schochetman G., Berkelman R.L.,  
 RA Economou A.N., Witte J.J., Furman L.J., Satten G.A., MacInnes K.A.,  
 RA Curran J.W., Jaffe H.W.,  
 RT "Molecular epidemiology of HIV transmission in a dental practice."  
 RL Science 256:1165-1171(1992).  
 DR EMBL: M92120; AAA44490.1;  
 FT NON\_TER 1 1  
 FT 16 16  
 SQ SEQUENCE 16 AA; 1745 MW; 72260DC174FF6428 CRC64;

Query Match 2.8%; Score 5; DB 12; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 7.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 NNTGD 161  
 ID 6 NNTGD 10

RESULT 3  
 ID 09QW64 PRELIMINARY; PRT; 18 AA.  
 AC 09QW64;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE SEZ-6 PROTEIN (FRAGMENT).  
 GN SEZ6 OR SEZ-6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALE/C;  
 RC MEDLINE=97225903; PubMed=9073173;  
 RA Herbst R., Nicklin M.J.H.;  
 RT "SEZ-6: promoter selectivity, genomic structure and localized  
 expression in the brain."  
 RL Mol. Brain Res. 44:309-322(1997).  
 DR EMBL: Y09922; CA71055.1;  
 DR MCD; MGI:104745; SEZ6.  
 FT NON\_TER 18 18  
 SQ SEQUENCE 18 AA; 1913 MW; F800A4350FFB7275 CRC64;

Query Match 2.8%; Score 5; DB 11; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 8.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AALL 21  
 ID 4 AALL 8

RESULT 4  
 ID 09R4A3 PRELIMINARY; PRT; 19 AA.  
 AC 09R4A3;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE N-ACETYL-D-NEURAMINIC ACID LYASE (EC 4.1.3.3) (FRAGMENT).  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=96276413; PubMed=8694758;  
 RA Ferrero M.A., Reglero A., Fernandez-Lopez M., Ordas R.,  
 RA Rodriguez-Aparicio L.B.;  
 RT "N-acetyl-D-neuraminic acid lyase generates the sialic acid for  
 colominic acid biosynthesis in Escherichia coli K1."  
 RL Biochem. J. 317:157-165(1996).  
 DR HSP; P06995; INAL.  
 SQ SEQUENCE 19 AA; 2061 MW; 72022247A20EBBE3 CRC64;

Query Match 2.8%; Score 5; DB 2; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AALL 21  
 ID 9 AALL 13

RESULT 5  
 ID 09TRV8 PRELIMINARY; PRT; 20 AA.  
 AC 09TRV8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE EIF-2 GAMMA.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=92002150; PubMed=1911855;  
 RA Bommer U.A., Kraft R., Kurzchalla T.V., Price N.T., Proud C.G.;  
 RT "Amino acid sequence analysis of the beta- and gamma-subunits of  
 eukaryotic initiation factor eIF-2. Identification of regions  
 interacting with GTP."  
 RL Biochim. Biophys. Acta 1079:308-315(1991).  
 SQ SEQUENCE 20 AA; 2036 MW; 419E28BA3857929E CRC64;

Query Match 2.8%; Score 5; DB 6; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 9.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 ALPEI 34  
 ID 10 ALPEI 14

RESULT 6  
 ID 069381 PRELIMINARY; PRT; 20 AA.  
 AC 069381;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE (HSV) DNA FOR IE mRNA-2 5' TERMINUS (FRAGMENT).  
 OS Herpes simplex virus (type 2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10310;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=84015370; PubMed=6312416;  
 RA Whitton J.L., Rixon F.J., Easton A.J., Clements J.B.;  
 RT "Immediate-early mRNA-2 of herpes simplex viruses types 1 and 2 is  
 RT unspliced: conserved sequences around the 5' and 3' termini correspond  
 RT to transcription regulatory signals.";  
 RL Nucleic Acids Res. 11:6271-6287(1983).  
 DR EMBL: X01624; CAA25766.1; -;  
 FT NON\_TER 20  
 SQ SEQUENCE 20 AA; 2180 MW; 11C39C0D07A447A9 CRC64;

Query Match  
 Best Local Similarity 2.8%; Score 5; DB 12; Length 20;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 LGIDL 77  
 DB 11 LGIDL 15

RESULT 7  
 O9TRC4  
 ID O9TRC4; PRELIMINARY; PRT; 22 AA.  
 AC O9TRC4;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE CYTOCHROME P450 DPB-3 ISOFORM (FRAGMENT).  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=94198820; PubMed=8185738;  
 RA Shiraga T., Iwasaki K., Nozaki K., Tamura T., Yamazoe Y., Kato R.,  
 RA Takahata A.;  
 RT "Isolation and characterization of four cytochrome P450 isozymes from  
 RT untreated and phenobarbital-treated beagle dogs.";  
 RL Biol. Pharm. Bull. 17:22-28(1994).  
 SQ SEQUENCE 22 AA; 2308 MW; 66F42EAD31280435 CRC64;

Query Match  
 Best Local Similarity 2.8%; Score 5; DB 6; Length 22;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 LLLAL 23  
 DB 7 LLLAL 11

RESULT 8  
 O68988  
 ID O68988; PRELIMINARY; PRT; 22 AA.  
 AC O68988;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)  
 DE (HSV) DNA FOR IE mRNA-2 5' TERMINUS (FRAGMENT).  
 OS human herpesvirus 1.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10298;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84015370; PubMed=6312416;  
 RA Whitton J.L., Rixon F.J., Easton A.J., Clements J.B.;  
 RT "Immediate-early mRNA-2 of herpes simplex viruses types 1 and 2 is  
 RT unspliced: conserved sequences around the 5' and 3' termini correspond  
 RT to transcription regulatory signals.";  
 RL Nucleic Acids Res. 11:6271-6287(1983).  
 DR EMBL: X01625; CAA25767.1; -;

FT NON\_TER 22  
 SQ SEQUENCE 22 AA; 2411 MW; 03EBE1C22C0D07A4 CRC64;

Query Match  
 Best Local Similarity 2.8%; Score 5; DB 12; Length 22;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 LGIDL 77  
 DB 11 LGIDL 15

RESULT 9  
 O87084  
 ID O87084; PRELIMINARY; PRT; 22 AA.  
 AC O87084;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE GLYCOPROTEIN C PRECURSOR (FRAGMENT).  
 GN GC.  
 OS Pseudorabies virus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Alphaherpesvirinae; Varicelloviruses.  
 OX NCBI\_TaxID=10345;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BECKER;  
 RX MEDLINE=96070924; PubMed=7499261;  
 RA Ryan P., Edwards C.O.;  
 RT "Systematic introduction of proline in a eukaryotic signal sequence  
 RT suggests asymmetry within the hydrophobic core.";  
 RL J. Biol. Chem. 270:27876-27879(1995).  
 DR EMBL: U29127; AAC54537.1; -;  
 KW Signal.  
 FT SIGNAL 1  
 FT NON\_TER 22  
 SQ SEQUENCE 22 AA; 2174 MW; A39469F5C3963F84 CRC64;

Query Match  
 Best Local Similarity 2.8%; Score 5; DB 12; Length 22;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 LLLAL 23  
 DB 9 LLLAL 13

RESULT 10  
 O04274  
 ID O04274; PRELIMINARY; PRT; 23 AA.  
 AC O04274;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE VIR-REPPRESSED PROTEIN (FRAGMENT).  
 GN VIR-18.  
 OC Bordetella pertussis.  
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Bordetella.  
 OX NCBI\_TaxID=520;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=18323;  
 RX MEDLINE=91072251; PubMed=2174866;  
 RA Beattie D.T., Knapp S., Mekalanos J.J.;  
 RT "Evidence that modulation requires sequences downstream of the  
 RT promoters of two vir-repressed genes of Bordetella pertussis.";  
 RL J. Bacteriol. 172:6997-7004(1990).  
 DR EMBL: M37228; AAA62718.1; -;  
 FT NON\_TER 23

SEQUENCE 23 AA; 2267 MW; DDD4E2140148C8BD CRC64;

Query Match 2.8%; Score 5; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AALL 20  
DB 12 AALL 16

RESULT 11

O9RSK9 PRELIMINARY; PRT; 23 AA.  
AC O9RSK9.  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE VACUOLATING TOXIN.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group.  
OC Helicobacter.  
OX NCBI\_TaxID=210;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=92268100; PubMed=1587837;  
RA Cover T.L., Blaser M.J.;  
RT "Purification and characterization of the vacuolating toxin from  
RT Helicobacter pylori."  
RT J. Biol. Chem. 267:10570-10575(1992).  
SQ SEQUENCE 23 AA; 2178 MW; DA91825A16071C0 CRC64;

Query Match 2.8%; Score 5; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 94 TTVII 98  
DB 4 TTVII 8

RESULT 12

O62594 PRELIMINARY; PRT; 23 AA.  
ID O62594.  
AC O62594.  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE NICOTINIC ACETYLCHOLINE RECEPTOR (FRAGMENT).  
GN NACHR.  
OS Rattus leucopus (mottie-tailed rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10115;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RX MEDLINE=94308229; PubMed=8034713;  
RA Walke W., Staple J., Adams L., Guey M., Chahine K., Goldman D.,  
RT "Calcium-dependent regulation of rat and chick muscle nicotinic  
RT acetylcholine receptor (nAChR) gene expression."  
RL J. Biol. Chem. 269:19447-19456(1994).  
DR EMBL; L19594; AAA21730.1; -.  
FT NON\_TER 23  
SQ SEQUENCE 23 AA; 2478 MW; 7EF72A3E7D4ACE45 CRC64;

Query Match 2.8%; Score 5; DB 11; Length 23;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 LLLAL 23  
DB 10 LLLAL 14

RESULT 13  
O9T2J1 PRELIMINARY; PRT; 24 AA.

AC O9T2J1.  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE 9 KDA PHOTOSYSTEM I PSII PROTEIN (FRAGMENT).  
OS Nicotiana tabacum (Common tobacco).  
OS Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; eusterids I; Solanales; Solanaceae; Nicotiana.  
OX NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=94105345; PubMed=8278548;  
RA Obokata J., Mikami K., Hayashida N., Nakamura M., Sugiyama M.;  
RT "Molecular heterogeneity of photosystem I. psal, psae, psaf, psah, and  
RT psal are all present in isoforms in Nicotiana spp."  
RT Plant Physiol. 102:1259-1267(1993).  
SQ SEQUENCE 24 AA; 2763 MW; 9D3398CE7C4531F2 CRC64;

Query Match 2.8%; Score 5; DB 8; Length 24;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 DLANT 112  
DB 13 DLANT 17

RESULT 14  
O9UL44 PRELIMINARY; PRT; 25 AA.

AC O9UL44.  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE LYSYL HYDROXYLASE 1 (FRAGMENT).  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Thai S.F., Yeowell H.N., Pinnett S.R.;  
RT "Characterization of the human lysyl hydroxylase promoter region."  
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF081786; AAD52693.1; -.  
FT NON\_TER 25  
SQ SEQUENCE 25 AA; 2732 MW; 12BD9D7AADC2AEBB CRC64;

Query Match 2.8%; Score 5; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 LLLAL 23  
DB 5 LLLAL 9

RESULT 15  
O99BS2 PRELIMINARY; PRT; 25 AA.  
ID O99BS2.  
AC O99BS2;

DT 01-JUN-2001 (TReMBLrel. 17, Created)  
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE REV PROTEIN (FRAGMENT).  
CN REV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TV013-B;  
RA Scriba T.J., Treurnicht F.K., Zeier M., Engelbrecht S.,  
RT "Characterization and phylogenetic analysis of South African HIV-1  
RT subtype C accessory genes";  
RL AIDS Res Hum. Retroviruses 0:0-0(2001).  
DR EMBL; AF325745; AAK09115.1; -.  
FT NON\_TER 25  
SQ SEQUENCE 25 AA; 2708 MW; D7933FE774F225A CRC64;

Query Match 2.88; Score 5; DB 12; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 18 ALIA 22  
| | | | |  
Db 11 ALIA 15

Search completed: March 28, 2002, 09:25:03  
Job time: 546 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using SW model

Run on: March 28, 2002, 09:25:03 ; Search time 56.59 Seconds  
(without alignments) updates/sec

US-09-726-348-2\_COPY\_126\_177

```

Perfect score: 52
Sequence: 1 QHVNCPPGGINAMNTITSYID.....NTGDEMCPEENGSCVDPDGG 52

```

Scoring table:  
OLIGO 60.0 , Gapext 60.0  
Gapop 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

```

Minimum DB seq length: 0
Maximum DB seq length: 30

```

post-processing: Listing first 45 summaries

```
Database :
1: sp.embryl.17.*
2: sp.bacteria.*
3: sp.fungi.*
4: sp.human.*
5: sp.invertebrate.*
6: sp.mammal.*
7: sp.mhc.*
8: sp.organelle.*
9: sp.phage.*
10: sp.plant.*
11: sp.rodent.*
12: sp.virus.*
13: sp.vertebrate.*
14: sp.unclassified.*
```

## SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.		No.					
1	5	9	6	16	12	078377	078377 human immun
2	5	9	6	27	22	0956D8	0956D8 escherichia
3	4	7	7	15	5	026323	026323 lymaea sta
4	4	7	7	16	2	P95712	P95712 shigella bo
5	4	7	7	16	4	09UR15	09UR15 homo sapien
6	4	7	7	16	4	09UC75	09UC75 homo sapien
7	4	7	7	16	5	026322	026322 lymaea sta
8	4	7	7	17	4	09UC28	09UC28 homo sapien
9	4	7	7	18	12	078375	078375 human immun
10	4	7	7	18	12	073920	073920 human immun
11	4	7	7	18	12	073921	073921 human immun
12	4	7	7	19	5	026321	026321 lymaea sta
13	4	7	7	19	12	078351	078351 human immun
14	4	7	7	20	4	09UC06	09UC06 homo sapien
15	4	7	7	20	12	073917	073917 human immun
16	4	7	7	20	12	073918	073918 human immun
17	4	7	7	21	2	09RAU0	09RAU0 streptomyce
18	4	7	7	21	8	034012	034012 beta webbia
19	4	7	7	22	10	095897	095897 citrus sine

20	4	7.7	23	11	p97918
21	4	7.7	24	10	Q47541
22	4	7.7	24	10	Q9AUF3
23	4	7.7	24	10	Q9AUF2
24	4	7.7	24	10	Q9AUF1
25	4	7.7	25	2	Q9X4T8
26	4	7.7	25	2	Q9S6D6
27	4	7.7	25	6	Q9TRY2
28	4	7.7	26	2	Q9L3B4
29	4	7.7	26	12	Q9CF14
30	4	7.7	27	2	Q08674
31	4	7.7	27	5	Q23745
32	4	7.7	27	5	Q06875
33	4	7.7	27	12	Q78325
34	4	7.7	27	12	Q78347
35	4	7.7	28	5	Q26318
36	4	7.7	28	5	Q26319
37	4	7.7	28	5	Q27356
38	4	7.7	29	2	Q9S5M6
39	4	7.7	29	12	Q78394
40	4	7.7	30	2	Q9R4J2
41	4	7.7	30	2	Q9R5J0
42	4	7.7	30	4	P78542
43	4	7.7	30	5	Q26320
44	4	7.7	30	5	Q78407
45	4	7.7	30	12	Q78407

## ALIGNMENTS

[illegible]

ID Q956D8 PRELIMINARY; PRT; 27 AA.  
 AC Q956D8;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
 DE SIMILAR TO KLEBSIELLA PNEUMONIAE ORF3 (FRAGMENT).  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=E75;  
 RX MEDLINE=99194747; PubMed=10094716;  
 RA Rahn A., Drummelsmith J., Whitfield C.;  
 RT "Conserved organization in the cps gene clusters for expression of  
 RT biosynthesis locus and the cps genes from Klebsiella pneumoniae.";  
 RL J. Bacteriol. 181:2307-2313(1999).  
 DR EMBL: AF118253; AAD30013.1; -  
 FT NON\_TER 27  
 SQ SEQUENCE 27 AA; 2859 MW; DE8DF51B1E841DCD CRC64;

Query Match  
 Best Local Similarity 9.6%; Score 5; DB 2; Length 27;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 ENGSC 46  
 Db 22 ENGSC 26

RESULT 3  
 ID Q26323 PRELIMINARY; PRT; 15 AA.  
 AC Q26323;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DE VDI/RPD2 ALPHA PEPTIDE (FRAGMENT).  
 OS Lymnaea stagnalis (Great pond snail).  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
 OC Lymnaeidae; Lymnaea.  
 NCBI\_TaxID=6523;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93351194; PubMed=8348590;  
 RA Bogard J., Van Kesteren R.E., Van Heerikhuizen H., Geraerts W.P.,  
 RA Veenstra J., Smit A.B., Joosse J.;  
 RT "Alternative splicing generates diversity of VDI/RPD2 alpha peptides  
 RT in the central nervous system of Lymnaea stagnalis.";  
 RL Cell. Mol. Neurobiol. 13:123-136(1993).  
 DR EMBL: S65078; AAB27755.1; -  
 FT NON\_TER 1  
 SQ SEQUENCE 15 AA; 1812 MW; 3F7AE460A6713943 CRC64;

Query Match  
 Best Local Similarity 7.7%; Score 4; DB 5; Length 15;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NCPG 7  
 Db 10 NCPG 13

RESULT 4  
 ID P95712 PRELIMINARY; PRT; 16 AA.  
 AC P95712;  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)

DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)  
 DE HYPOTHETICAL PROTEIN IN FILE 5' REGION (ORF8) (FRAGMENT).  
 OS Shigella boydii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Shigella.  
 NCBI\_TaxID=621;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC9733;  
 RX MEDLINE=97369817; PubMed=9226258;  
 RA Al Mamun A.A.M., Tomimaga A., Enomoto M.;  
 RL J. Bacteriol. 179:4493-4500(1997).  
 DR EMBL: D89825; BAA14025.1; -  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 16 AA; 1804 MW; 16C7694385F1E377 CRC64;

Query Match  
 Best Local Similarity 7.7%; Score 4; DB 2; Length 16;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GGIN 10  
 Db 11 GGIN 14

RESULT 5  
 ID Q90B15 PRELIMINARY; PRT; 16 AA.  
 AC Q90B15;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DE CYSTATIN S.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=92138674; PubMed=1778989;  
 RA Isemura S., Saitoh E., Sanada K., Minakata K.;  
 RT "Identification of full-sized forms of salivary (S-type) cystatins  
 RT (cystatin SM, cystatin SA, cystatin S, and two phosphorylated forms of  
 RT cystatin S) in human whole saliva and determination of phosphorylation  
 RL sites of cystatin S.";  
 RL J. Biochem. 110:648-654(1991).  
 RN [2]  
 RP SEQUENCE.  
 RX MEDLINE=92074898; PubMed=1741693;  
 RA Johnson M., Richardson C.F., Bergey E.J., Levine M.J.,  
 RA Nancollis G.H.;  
 RT "The effects of human salivary cystatins and statherin on  
 RT hydroxyapatite crystallization.";  
 RL Arch. Oral Biol. 36:631-636(1991)  
 SQ SEQUENCE 16 AA; 1765 MW; A6EDF31C3186F9C9 CRC64;

Query Match  
 Best Local Similarity 7.7%; Score 4; DB 4; Length 16;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 PGCI 9  
 Db 11 PGCI 14

RESULT 6  
 ID Q90C75 PRELIMINARY; PRT; 16 AA.  
 AC Q90C75;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)



DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
 DE 01-JUN-2000 (TRENBLREL. 14, Last annotation update)  
 DE DIPEPTIDYLPEPTIDASE IV (EC 3.4.14.5) (FRAGMENT).  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=95294018; PubMed=7539799;  
 RA Duke-Cohan J.S., Morimoto C., Rucker J.A., Schlossman S.F.;  
 RT "A novel form of dipeptidylpeptidase IV found in human serum.  
 RT Isolation, characterization, and comparison with T lymphocyte membrane  
 RT dipeptidylpeptidase IV (CD26).";  
 RL J. Biol. Chem. 270:14107-14114(1995).  
 SO SEQUENCE 16 AA; 1600 MW; DFB3DD95FE0A1EB CRC64;

Query Match 7.7%; Score 4; DB 4; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 DPG 52  
 DB 10 DPG 13

RESULT 7  
 ID 026322 PRELIMINARY; PRT; 16 AA.  
 AC 026322;  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DE VDI/RPD2 ALPHA PEPTIDE (FRAGMENT).  
 OS Lymanaea stagnalis (Great pond snail).  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
 OC Lymanaeidae; Lymanaea.  
 NCBI\_TaxID=6523;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93351194; PubMed=8348590;  
 RA Bogerd J., Van Kesteren R.E., van Heerikhuizen H., Geraerts W.P.,  
 RA Veensstra J., Smit A.B., Joosse J.;  
 RT "Alternative splicing generates diversity of VDI/RPD2 alpha peptides  
 RT in the central nervous system of Lymanaea stagnalis.";  
 RL Cell. Mol. Neurobiol. 13:123-136(1993).  
 DR EMBL: S65072; AAB2754.1; -;  
 FT NON\_TER 1 1  
 SO SEQUENCE 16 AA; 1869 MW; 4EBAE460A67907A CRC64;

Query Match 7.7%; Score 4; DB 5; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 NCPG 7  
 DB 11 NCPG 14

RESULT 8  
 ID 09UC28 PRELIMINARY; PRT; 17 AA.  
 AC 09UC28;  
 DT 01-MAY-2000 (TRENBLREL. 13, Created)  
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)  
 DE 85 KDA GLIOMA MEMBRANE PROTEIN/CD44 HOMOLOG.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=95079448; PubMed=7527301;  
 RA Okada H., Yoshida J., Seo H., Wakabayashi T., Sugita K., Hagiwara M.;  
 RT "Anti-(glioma surface antigen) monoclonal antibody G-22 recognizes  
 RT overexpressed CD44 in glioma cells.";  
 RL Cancer Immunol. Immunother. 39:313-317(1994).  
 SO SEQUENCE 17 AA; 1788 MW; 7806F51C97A3E217 CRC64;

Query Match 7.7%; Score 4; DB 4; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 NNTG 35  
 DB 9 NNTG 12

RESULT 9  
 ID 078375 PRELIMINARY; PRT; 18 AA.  
 AC 078375;  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DE VIRAL SAMPLE FLUOR4B (FLORIDA PATIENT B), PARTIAL ENV CDS, V4 REGION  
 DE (FRAGMENT).  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zhang L.O., Leigh-Brown A.J.;  
 RL Submitted (Apr-1992) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92271245; PubMed=1589796;  
 RA Ou C.Y., Ciesielski C.A., Myers G., Bandea C.I., Luo C.C.,  
 RA Korber B.T.M., Mullins J.I., Schochetman G., Berkelman R.L.,  
 RA Economou A.N., Witte J.J., Furman L.J., Satten G.A., MacInnes K.A.,  
 RA Curran J.W., Jaffe H.W.;  
 RT "Molecular epidemiology of HIV transmission in a dental practice.";  
 RL Science 256: 165-171(1992).  
 DR EMBL: M62119; AAA44489.1; -;  
 FT NON\_TER 1 1  
 SO SEQUENCE 18 AA; 1918 MW; 005E4797B96E0006 CRC64;

Query Match 7.7%; Score 4; DB 12; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 NNTG 35  
 DB 8 NNTG 11

RESULT 10  
 ID 073920 PRELIMINARY; PRT; 18 AA.  
 AC 073920;  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DE ENVELOPE GLYCOPROTEIN, V1-V2 REGION (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-PATIENT 6;  
 MEDLINE-96251940; PubMed-8661394;  
 RA Palmer C., Balfe P., Fox D., May J.C., Frederiksson R., Fenyo E.M.,  
 RT McKeating J.A.;  
 RL "Functional characterization of the V1V2 region of human  
 immunodeficiency virus type 1.";  
 RL Virology 220:436-449(1996).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-PATIENT 6;  
 RA Blouin C.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U52338; AAB07992.1;  
 KM Envelope protein.  
 FT NON-TER  
 SQ SEQUENCE 18 AA; 2067 MW; 67B44AD5F6920C78 CRC64;

Query Match 7.7%; Score 4; DB 12; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 NTIT 16  
 DB 12 NTIT 15

RESULT 11  
 073921 PRELIMINARY; PRT; 18 AA.  
 AC 073921;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)  
 DE ENVELOPE GLYCOPROTEIN, V1-V2 REGION (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_Taxid=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PATIENT 6;  
 RX MEDLINE-96251940; PubMed-8661394;  
 RA Palmer C., Balfe P., Fox D., May J.C., Frederiksson R., Fenyo E.M.,  
 RT McKeating J.A.;  
 RL "Functional characterization of the V1V2 region of human  
 immunodeficiency virus type 1.";  
 RL Virology 220:436-449(1996).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-PATIENT 6;  
 RA Blouin C.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U52339; AAB07993.1;  
 KM Envelope protein.  
 FT NON-TER  
 SQ SEQUENCE 18 AA; 2067 MW; 67B44AD5F6920C78 CRC64;

Query Match 7.7%; Score 4; DB 12; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 NTIT 16  
 DB 12 NTIT 15

RESULT 12  
 026321 PRELIMINARY; PRT; 19 AA.  
 AC 026321;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)  
 DE VDI/RPD2 ALPHA PEPTIDE (FRAGMENT).  
 OS Lymnaea stagnalis (Great pond snail).  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
 OC Lymnaeidae; Lymnaea.  
 OX NCBI\_Taxid=6523;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-93351194; PubMed-8348590;  
 RA Bogerd J., Van Kesteren R.E., van Heerikhuisen H., Geraerts W.P.,  
 RA Veenstra J., Smit A.B., Joosse J.;  
 RT "Alternative splicing generates diversity of VDI/RPD2 alpha peptides  
 in the central nervous system of Lymnaea stagnalis.";  
 RL Cell. Mol. Neurobiol. 13:123-136(1993).  
 DR EMBL: S65071; AAB27753.1;  
 FT NON-TER  
 SQ SEQUENCE 19 AA; 2110 MW; 069874E166799F4C CRC64;

Query Match 7.7%; Score 4; DB 5; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 NCPG 7  
 DB 14 NCPG 17

RESULT 13  
 078351 PRELIMINARY; PRT; 19 AA.  
 AC 078351;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)  
 DE VIRAL SAMPLE FLIBRA (FLORIDA PATIENT B), PARTIAL ENV CDS, V4 REGION  
 DE (FRAGMENT).  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_Taxid=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zhang L.Q., Leigh-Brown A.J.;  
 RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE-92271245; PubMed-1589796;  
 RA Ou C.Y., Ciecielski C.A., Myers G., Banea C.I., Luo C.C.,  
 RA Korber B.T.M., Mullins J.I., Schochetman G., Berkman R.L.,  
 RA Economou A.N., Witte J.J., Furman L.J., Satten G.A., MacInnes K.A.,  
 RA Curran J.W., Jaffe H.W.;  
 RT "Molecular epidemiology of HIV transmission in a dental practice.";  
 RL Science 256:1165-1171(1992).  
 DR EMBL: M92118; AAA44488.1;  
 FT NON-TER  
 SQ SEQUENCE 19 AA; 1965 MW; 5FEC8787D98CEA1C CRC64;

Query Match 7.7%; Score 4; DB 12; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 NNTG 35  
 DB 8 NNTG 11

RESULT 14  
 090C06 PRELIMINARY; PRT; 20 AA.  
 ID 090C06  
 AC 090C06;

Search completed: March 28, 2002, 09:25:05  
 Job time: 548 sec

DF 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)  
 DE CYSSTATIN SN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE.  
 RX MEDLINE=92138674; PubMed=1778989;  
 RA Isemura S., Saitoh E., Sanada K., Minakata K.;  
 RT "Identification of full-sized forms of salivary (S-type) cystatins  
 RT (cystatin SN, cystatin SA, cystatin S, and two phosphorylated forms of  
 RT cystatin S) in human whole saliva and determination of phosphorylation  
 RT sites of cystatin S.";  
 RL J. Biochem. 110:648-654(1991).  
 DR InterPro: IPR001005; Myb DNA bind.  
 DR PROSITE: PS00037; MYB.1; UNKNOWN.1.  
 SQ SEQUENCE 20 AA; 2287 MW; 482872B245C5511C CRC64;

Query Match 7.7%; Score 4; DB 4; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PGCI 9  
 ||||  
 Db 11 PGCI 14

RESULT 15  
 073917  
 ID 073917 PRELIMINARY; PRT; 20 AA.  
 AC 073917;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, last annotation update)  
 DE ENVELOPE GLYCOPROTEIN, VI-V2 REGION (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retrovirdae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PATIENT 6;  
 RX MEDLINE=96251940; PubMed=8661394;  
 RA Palmer C., Baile P., Fox D., May J.C., Frederiksson R., Fenyo E.M.,  
 RA McKeating J.A.;  
 RT "Functional characterization of the V1V2 region of human  
 RT immunodeficiency virus type 1.";  
 RT Virology 220:436-449(1996).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PATIENT 6;  
 RA Blouin C.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U52335; AAB07989.1; -.  
 KW Envelope protein.  
 FT NON\_TER 1  
 FT 1  
 SQ SEQUENCE 20 AA; 2283 MW; 6B58C3B1AD5F692 CRC64;

Query Match 7.7%; Score 4; DB 12; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 NTIT 16  
 ||||  
 Db 12 NTIT 15

Thu Mar 28 09:21:37 2002

us-09-726-348-2\_copy\_126\_177.rspt

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: March 28, 2002, 09:15:57 ; Search time 56.59 Seconds  
(without alignments)  
591.913 Million cell updates/sec

Title: US-09-726-348-2

Perfect score: 229  
Sequence: 1 MAPHGPGSTTLVPMMAALL.....TTLVSILLMAGRKAKTS 229

Scoring table: OLIGO  
Gapop 60.0 , Gapect 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 12410

Minimum DB seq length: 0  
Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database :

SPREMBL\_17:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.protist:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2.6	22	11	Q9WUK1	Q9WUK1 ratu
2	2.6	23	11	Q9CT79	Q9CT79 mus
3	2.6	30	4	Q9UMJ2	Q9UMJ2 homo
4	2.2	15	2	Q9RS45	Q9RS45 myco
5	2.2	16	12	Q78377	Q78377 huma
6	2.2	18	11	Q9QW64	Q9QW64 mus
7	2.2	18	11	Q9CRO3	Q9CRO3 mus
8	2.2	19	2	Q9RA43	Q9RA43 esche
9	2.2	20	6	Q9TRV8	Q9TRV8 orycto
10	2.2	20	12	Q9G381	Q9G381 herpes
11	2.2	22	6	Q9TRC4	Q9TRC4 canis
12	2.2	22	12	Q68588	Q68588 huma
13	2.2	22	12	Q87084	Q87084 pseud
14	2.2	23	2	Q04274	Q04274 bordet
15	2.2	23	2	Q9RSK9	Q9RSK9 helicob
16	2.2	23	11	Q62594	Q62594 ratu
17	2.2	23	11	Q9Z2R1	Q9Z2R1 ratu
18	2.2	24	8	Q9T2J1	Q9T2J1 nicot
19	2.2	25	4	Q9UL44	Q9UL44 homo

20	5	2.2	25	12	Q9B852	Q9B852 human
21	5	2.2	25	12	Q9B8R8	Q9B8R8 human
22	5	2.2	26	2	Q9424	Q9424 pseudom
23	5	2.2	26	2	Q9NP34	Q9NP34 homo
24	5	2.2	26	13	Q9PS44	Q9PS44 gallu
25	5	2.2	26	13	Q9PS43	Q9PS43 gallu
26	5	2.2	26	13	Q9S6D8	Q9S6D8 esche
27	5	2.2	27	6	Q29019	Q29019 sus scro
28	5	2.2	27	6	Q38645	Q38645 bacteri
29	5	2.2	27	12	Q56495	Q56495 hepatit
30	5	2.2	27	12	Q9Y2Q5	Q9Y2Q5 hepatit
31	5	2.2	27	13	Q9PRY4	Q9PRY4 gallu
32	5	2.2	28	8	Q9ZY54	Q9ZY54 leishma
33	5	2.2	28	8	Q38060	Q38060 bacteri
34	5	2.2	28	11	Q62677	Q62677 ratu
35	5	2.2	30	2	Q44171	Q44171 anabaena
36	5	2.2	30	10	Q9S8U1	Q9S8U1 chlamydo
37	5	1.7	8	11	Q9ET18	Q9ET18 mus spre
38	4	1.7	8	11	Q9ET17	Q9ET17 mus caroli
39	4	1.7	8	11	Q9ET16	Q9ET16 mesocricetu
40	4	1.7	9	7	Q9MW43	Q9MW43 homo sapien
41	4	1.7	9	8	Q9TKD9	Q9TKD9 pericalymma
42	4	1.7	9	13	P82075	P82075 ittoria
43	4	1.7	9	13	P82093	P82093 ittoria
44	4	1.7	10	4	Q9H3R9	Q9H3R9 homo sapien
45	4	1.7	10	9	Q9TOR6	Q9TOR6 bacterioph

## ALIGNMENTS

RESULT 1  
ID Q9WUK1 PRELIMINARY; PRT; 22 AA.

AC Q9WUK1;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE RELAXIN-LIKE FACTOR (FRAGMENT).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY;  
RA Spies A.-N., Fusch W., Ivell R.;  
RT "Cloning and sequencing of the rat relaxin-like factor and its  
RT promoter."  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF139920; AAD33851.1; -  
FT NON\_TER  
FT SEQUENCE 22 AA; 2325 MW; F1AA6B6118F143F CXC64;

Query Match 2.6%; Score 6; DB 11; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LLLALG 24  
| | | | |  
Db 8 LLLALG 13

RESULT 2  
ID Q9CT79 PRELIMINARY; PRT; 23 AA.  
AC Q9CT79;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 1110068LDIRK PROTEIN (FRAGMENT).  
GN 1110068LDIRK

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RM MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamazaki I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RT Nature 409:685-690(2001).  
 RL EMBL: AK004410; BAB3235.1.  
 DR MGD: MGI:1916116; 111008L01R1K.  
 FT NON\_TER 1  
 SQ SEQUENCE 23 AA; 2396 MW; 4DC8245449828A0A CRC64;

Query Match 2.6%; Score 6; DB 11; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 197 SFSLLM 202  
 DB 13 SFSLLM 18

RESULT 3  
 Q9UMJ2 PRELIMINARY; PRT; 30 AA.  
 AC Q9UMJ2:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE IFNA PROTEIN PRECURSOR (FRAGMENT).  
 GN IFNA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RM MEDLINE=85155470; PubMed=63974400;  
 RA Fuke M., Hendrix L.C., Bollon A.P.;  
 RT "Pseudogene IFN-alpha L: removal of the stop codon in the signal  
 RT sequence permits expression of active human interferon.";  
 RL Gene 32:135-140(1984).  
 CC -1 SIMILARITY: BELONGS TO THE INTERFERON ALPHA, BETA AND DELTA  
 CC FAMILY.  
 DR EMBL: M10201; AAA52732.1; -;  
 DR InterPro: IPR000471; Interferon\_abd.  
 DR Pfam: PF00143; Interferon; 1.  
 KW Antiviral; Cytokine; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 >30 POTENTIAL.  
 FT NON\_TER 30  
 SQ SEQUENCE 30 AA; 3242 MW; 93A5E56C69257C3B CRC64;

Query Match 2.6%; Score 6; DB 4; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 197 SFSLLM 202  
 DB 4 SFSLLM 9  
 RESULT 4  
 Q9R545 PRELIMINARY; PRT; 15 AA.  
 AC Q9R545:  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE 30S RIBOSOMAL PROTEIN (FRAGMENT).  
 OS Mycobacterium bovis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE.  
 RM MEDLINE=94009653; PubMed=8405418;  
 RA Ohara N., Kimura M., Higashi Y., Yamada T.;  
 RT "Isolation and amino acid sequence of the 30S ribosomal protein S19  
 RT from Mycobacterium bovis BCG.";  
 RL FBS Lett. 331:9-14(1993).  
 FT NON\_TER 1  
 SQ SEQUENCE 15 AA; 1674 MW; 07A36F018AE355A4 CRC64;

Query Match 2.2%; Score 5; DB 2; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 223 RRRKAK 227  
 DB 10 RRRKAK 14

RESULT 5  
 Q78377 PRELIMINARY; PRT; 16 AA.  
 AC Q78377:  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE VIRAL SAMPLE FLIPBR4C (FLORIDA PATIENT B), PARTIAL ENV CDS, V4 REGION  
 DE (FRAGMENT).  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroviruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RM Zhang L.O., Leigh-Brown A.J.;  
 RL Submitted (Apr-1992) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RM MEDLINE=92271245; PubMed=1589796;  
 RX Ou C.Y., Ciesielski C.A., Myers G., Banda C.I., Luo C.C.,  
 RA Korber B.T.M., Mullins J.I., Schochetman G., Berkelman R.L.,  
 RA Economou A.N., Witte J.J., Furman L.J., Satten G.A., MacInnes K.A.,  
 RA Curran J.W., Jaffe H.W.;  
 RT "Molecular epidemiology of HIV transmission in a dental practice.";  
 RL Science 256:1165-1171(1992).  
 DR EMBL: M92120; AAA44490.1; -;  
 FT NON\_TER 1  
 FT CHAIN 16  
 FT NON\_TER 16  
 SQ SEQUENCE 16 AA; 1745 MW; 72260DC174FF6428 CRC64;

Query Match 2.2%; Score 5; DB 12; Length 16;  
Best Local Similarity 100.0%; Pred. No. 9.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 157 NNTGD 161  
DB 6 NNTGD 10

RESULT 6  
ID 09QMG4 PRELIMINARY; PRT; 18 AA.  
AC 09QMG4;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE SEZ-6 PROTEIN (FRAGMENT).  
CN SEZ-6 OR SEZ-6.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPRAIN-BALB/C;  
RX MEDLINE=97225903; PubMed=9073173;  
RA Herbst R., Nicklin M.J.H.;  
RT "SEZ-6: promoter selectivity, genomic structure and localized  
expression in the brain.";  
RL MOL. Brain Res. 44:309-322(1997).  
DR EMBL: Y09922; CAA71055.1; -.  
DR MGD: MGI:104745; Sez6.  
FT NON\_TER  
SQ SEQUENCE 18 AA; 1913 MW; F800A4350FEB7275 CRC64;

Query Match 2.2%; Score 5; DB 11; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AALLL 21  
DB 4 AALLL 8  
RESULT 7  
ID 09CRO3 PRELIMINARY; PRT; 18 AA.  
AC 09CRO3;  
DT 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE 170012913RIK PROTEIN (FRAGMENT).  
GN 170012913RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LUNG;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Arakawa T., Hara A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide T., Pesole G., Quackenbush J.,  
RA Schmitt L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Botfield D., Bonjuna N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Guslinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL: AK018461; BAB31224.1; -.  
DR MGD: MGI:1915183; 170012913RIK.  
FT NON\_TER  
SQ SEQUENCE 18 AA; 2038 MW; 65C427525FEFBA1 CRC64;

Query Match 2.2%; Score 5; DB 11; Length 18;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 220 ATQRR 224  
DB 2 ATQRR 6

RESULT 8  
ID 09R4A3 PRELIMINARY; PRT; 19 AA.  
AC 09R4A3;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
DE N-ACETYL-D-NEURAMINIC ACID LYASE (EC 4.1.3.3) (FRAGMENT).  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=96276413; PubMed=8694758;  
RA Ferrero M.A., Regliero A., Fernandez-Lopez M., Ordas R.,  
RA Rodriguez-Aparicio L.B.;  
RT "N-acetyl-D-neuraminic acid lyase generates the sialic acid for  
RT colominic acid biosynthesis in Escherichia coli K1.";  
RL Biochem. J. 317:157-165(1996).  
DR HSSP: P06995; INAL.  
SQ SEQUENCE 19 AA; 2061 MW; 72022247A20EBBE3 CRC64;

Query Match 2.2%; Score 5; DB 2; Length 19;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AALLL 21  
DB 9 AALLL 13

RESULT 9  
ID 09TRV8 PRELIMINARY; PRT; 20 AA.  
AC 09TRV8;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
DE EIF-2 GAMMA.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=92002150; PubMed=1911855;  
RA Bommer U.A., Kraft R., Kurzchalia T.V., Price N.T., Proud C.G.;

RT "Amino acid sequence analysis of the beta- and gamma-subunits of  
RT eukaryotic initiation factor eIF-2. Identification of regions  
RT interacting with GTP." 1079:308-315(1991)  
RT Biochim. Biophys. Acta 1079:308-315(1991)  
SQ SEQUENCE 20 AA; 2036 MW; 419E2B8A3857929E CRC64;

Query Match 2.2%; Score 5; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ALPEI 34  
Db 10 ALPEI 14

RESULT 10  
Q69381 PRELIMINARY; PRT; 20 AA.  
AC 069381;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)  
DE (HSV) DNA FOR IE MRNA-2 5' TERMINUS (FRAGMENT).  
OS Herpes simplex virus (type 2).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
NCBI\_TaxID=10310;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84015370; PubMed=6312416;  
RA Whitton J.L., Rixon F.J., Easton A.J., Clements J.B.;  
RT "Immediate-early mRNA-2 of herpes simplex viruses types 1 and 2 is  
RT to transcription regulatory signals.";  
RL Nucleic Acids Res. 11:6271-6287(1983).  
DR EMBL: X01624; CAA25766.1; -.  
FT NON\_TER 20  
SQ SEQUENCE 20 AA; 2180 MW; 11C39C0D07A47A9 CRC64;

Query Match 2.2%; Score 5; DB 12; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 LGIDL 77  
Db 11 LGIDL 15

RESULT 11  
Q9TRC4 PRELIMINARY; PRT; 22 AA.  
AC 09TRC4;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
DE CYTOCHROME P450 DBP-3 ISOFORM (FRAGMENT).  
OS Canis familiaris (dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=94198820; PubMed=8185738;  
RA Shiraga T., Iwasaki K., Nozaki K., Tamura T., Yamazoe Y., Kato R.,  
RA Takanaoka A.;  
RT "Isolation and characterization of four cytochrome P450 isozymes from  
RT untreated and phenobarbital-treated beagle dogs.";  
RL Biol. Pharm. Bull. 17:22-28(1994).  
SQ SEQUENCE 22 AA; 2308 MW; 66F42EAD31280435 CRC64;

Query Match 2.2%; Score 5; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LILAL 23  
Db 7 LILAL 11

RESULT 12  
Q68988 PRELIMINARY; PRT; 22 AA.  
AC 068988;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)  
DE (HSV) DNA FOR IE MRNA-2 5' TERMINUS (FRAGMENT).  
OS human herpesvirus 1.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Simplexvirus.  
NCBI\_TaxID=10298;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=84015370; PubMed=6312416;  
RA Whitton J.L., Rixon F.J., Easton A.J., Clements J.B.;  
RT "Immediate-early mRNA-2 of herpes simplex viruses types 1 and 2 is  
RT to transcription regulatory signals.";  
RL Nucleic Acids Res. 11:6271-6287(1983).  
DR EMBL: X01625; CAA25767.1; -.  
FT NON\_TER 22  
SQ SEQUENCE 22 AA; 2411 MW; 03E8E1C22C0D07A4 CRC64;

Query Match 2.2%; Score 5; DB 12; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 LGIDL 77  
Db 11 LGIDL 15

RESULT 13  
Q87084 PRELIMINARY; PRT; 22 AA.  
AC 087084;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)  
DE GLYCOPROTEIN C PRECURSOR (FRAGMENT).  
GN GC.  
OS Pseudorabies virus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Alphaherpesvirinae; Varicellovirus.  
NCBI\_TaxID=10345;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BECKER;  
RC MEDLINE=96070924; PubMed=7499261;  
RA Ryan P., Edwards C.O.;  
RT "Systematic introduction of proline in a eukaryotic signal sequence  
RT suggests asymmetry within the hydrophobic core.";  
RL J. Biol. Chem. 270:27876-27879(1995).  
DR EMBL: U29127; AAC54537.1; -.  
FT SIGNAL 1  
FT NON\_TER 21  
SQ SEQUENCE 22 AA; 2174 MW; A39469F5C3963F84 CRC64;

Query Match 2.2%; Score 5; DB 12; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.2e+03;



Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 19 LLLAL 23  
 Db 9 LLLAL 13

Search completed: March 28, 2002, 09:25:03  
 Job time: 546 sec

RESULT 14  
 004274  
 ID 004274 PRELIMINARY; PRT; 23 AA.  
 AC 004274;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
 DE VIR-REPPRESSED PROTEIN (FRAGMENT).  
 GN VIR-18.  
 OS Bordetella pertussis.  
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
 OC Bordetella.  
 OX NCB1\_TaxID=520;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=18323;  
 RX MEDLINE=91072251; PubMed=2174866;  
 RA Beattie D.T., Knapp S., Mekalanos J.J.;  
 RT "Evidence that modulation requires sequences downstream of the  
 promoters of two vir-repressed genes of Bordetella pertussis.";  
 RL J. Bacteriol. 172:6997-7004(1990).  
 DR EMBL; M37228; AAA62718.1;  
 FT NON\_TER 23  
 SQ SEQUENCE 23 AA; 2267 MW; DDD4E2140148C8BD CRC64;

Query Match 2.2%; Score 5; DB 2; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAALL 20  
 Db 12 AAALL 16

RESULT 15  
 09R5K9  
 ID 09R5K9 PRELIMINARY; PRT; 23 AA.  
 AC 09R5K9;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
 DE VACUOLATING TOXIN.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCB1\_TaxID=210;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=92268100; PubMed=1587837;  
 RA Cover T.L., Blaser M.J.;  
 RT "Purification and characterization of the vacuolating toxin from  
 Helicobacter pylori.";  
 RL J. Biol. Chem. 267:10570-10575(1992).  
 SQ SEQUENCE 23 AA; 2178 MW; DA91825EA16071C0 CRC64;

Query Match 2.2%; Score 5; DB 2; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 TTVII 98  
 Db 4 TTVII 8

Thu Mar 28 09:21:31 2002

us-09-726-348-2.ispt

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:14:46 ; Search time 62.93 Seconds  
(without alignments)  
103.125 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_1\_177

Perfect score: 971  
Sequence: 1 MAPHGPGLTTLVPMALL.....NTGDEPKCPKNSCVDPDGG 177

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.5	11.8	2907	1 FBN2_MOUSE	Q61555 mus musculus
2	105.5	10.9	2911	1 FBN2_HUMAN	P35556 homo sapien
3	95.5	9.8	2871	1 FBN1_HUMAN	P35555 homo sapien
4	91	9.4	1114	1 TCPE8_CHICK	P36197 gallus galli
5	88.5	9.1	638	1 Y053_HUMAN	P43331 homo sapien
6	88.5	9.1	2871	1 FBN1_MOUSE	Q61554 mus musculus
7	86.5	8.9	2871	1 FBN1_BOVIN	P98133 bos taurus
8	82	8.4	769	1 AD11_HUMAN	O75078 homo sapien
9	82	8.4	773	1 AD11_MOUSE	Q921V4 mus musculus
10	81	8.3	432	1 FGL2_MOUSE	P12804 mus musculus
11	78.5	8.1	2871	1 FBN1_PIG	O91V36 sus scrofa
12	76.5	7.9	2213	1 SORL1_RABIT	O95209 o. sortilin-
13	75	7.6	886	1 EMRL_HUMAN	Q14246 homo sapien
14	74	7.6	638	1 KAL_HUMAN	P03852 homo sapien
15	73	7.5	979	1 PUT3_YEAST	P23502 saccharomyc
16	72.5	7.5	641	1 SCAB_RABIT	O97742 oryctolagus
17	72.5	7.5	1019	1 ENTK_HUMAN	P98073 homo sapien
18	72	7.4	597	1 CHTX_SPROI	Q05638 streptomyc
19	72	7.4	1043	1 TCF8_MESAU	Q60542 mesocricetu
20	72	7.4	2019	1 CINE_RAT	P15389 rattus norv
21	71.5	7.4	495	1 Y243_AOUAE	O66606 agulifex aeo
22	71.5	7.4	8759	1 MAMB_CAPII	O95327 capra hircu
23	71.5	7.4	1124	1 TCPE8_HUMAN	P33725 homo sapien
24	71.5	7.4	1173	1 TSPL_XENLA	P35448 xenopus lae
25	71.5	7.4	1895	1 YIK3_CAEEL	P41551 caenorhabd1
26	71	7.3	158	1 VE6_HPV45	P21735 human papil
27	71	7.3	253	1 KLT7_HUMAN	P49862 homo sapien
28	71	7.3	452	1 AD11_XENLA	O95223 xenopus lae
29	71	7.3	1091	1 JSN1_YEAST	P08135 saccharomyc
30	71	7.3	1951	1 C1N3_RAT	P08104 rattus norv
31	71	7.3	1964	1 NTC4_MOUSE	P31695 mus musculus
32	70.5	7.3	204	1 IP22_CAPAN	O49146 capicum an
33	70.5	7.3	377	1 PSPB_MOUSE	P50405 mus musculus

34	70.5	7.3	425	1 CEPL_HUMAN	Q92989 homo sapien
35	70.5	7.3	1367	1 VGR2_MOUSE	P35918 mus musculus
36	70.5	7.3	1592	1 SORL1_CHICK	Q98930 g. sortilin-
37	70	7.2	158	1 VE6_HPV18	P06463 human papil
38	70	7.2	279	1 COAT_SBMV	P03607 southern be
39	70	7.2	400	1 VMSA_HPBV2	P03141 hepatitis b
40	69.5	7.2	233	1 CYCX_NITED	Q50926 nitrosomona
41	69.5	7.2	335	1 OTCC_BACIL	O86132 bacillus li
42	69	7.1	199	1 CD69_MOUSE	P37217 mus musculus
43	69	7.1	484	1 LEM2_PIG	P98110 sus scrofa
44	69	7.1	712	1 FBL1_CAEEL	O77469 caenorhabd1
45	69	7.1	1097	1 S24C_AKATH	Q9m291 arabidopsis

## ALIGNMENTS

RESULT 1  
FBN2\_MOUSE STANDARD; PRT; 2907 AA.  
ID FBN2\_MOUSE  
AC Q61555; Q63957;  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE FIBRILLIN 2 PRECURSOR.  
GN FBN2 OR FBN-2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95263670; PubMed=7744963;  
RA Zhang H., Hu W., Ramirez F.;  
RT "Developmental expression of fibrillin genes suggests heterogeneity  
of extracellular microfibrils".  
RL J. Cell Biol. 129:1165-1176(1995).  
[2]  
RP SEQUENCE OF 210-317 FROM N.A.  
RX MEDLINE=94140368; PubMed=8307578;  
RA Li X., Pereira L., Zhang H., Sanguinetti C., Ramirez F., Bonadio J.,  
Francke U.;  
RT "Fibrillin genes map to regions of conserved mouse/human synteny on  
mouse chromosomes 2 and 18".  
RL Genomics 18:667-672(1993).  
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS  
THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE  
LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.  
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING  
EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: L39790; AAA74908.1; -;  
CC EMBL: S69359; AAC60685.1; -;  
DR MGD: MGI:95490; Fbn2.  
DR InterPro: IPR00152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF-Ca.  
DR InterPro: IPR001438; EGF-IL.  
DR InterPro: IPR002212; TB.  
DR InterPro: IPR000822; Znf-C2H2.  
DR Pfam: PF00008; EGF; 46.  
DR Pfam: PF00683; TB; 9.  
DR PRINTS: PR00010; EGFBL00D.  
DR SMART: SM00179; EGF\_CA; 43.

FT	DISOFLIFD	287	301	BY	SIMILARITY
FT	DISOFLIFD	303	316	BY	SIMILARITY
FT	DISOFLIFD	322	334	BY	SIMILARITY
FT	DISOFLIFD	329	343	BY	SIMILARITY
FT	DISOFLIFD	345	358	BY	SIMILARITY
FT	DISOFLIFD	491	503	BY	SIMILARITY
FT	DISOFLIFD	498	512	BY	SIMILARITY
FT	DISOFLIFD	514	526	BY	SIMILARITY
FT	DISOFLIFD	532	542	BY	SIMILARITY
FT	DISOFLIFD	537	551	BY	SIMILARITY
FT	DISOFLIFD	553	566	BY	SIMILARITY
FT	DISOFLIFD	572	584	BY	SIMILARITY
FT	DISOFLIFD	579	593	BY	SIMILARITY
FT	DISOFLIFD	595	608	BY	SIMILARITY
FT	DISOFLIFD	614	625	BY	SIMILARITY
FT	DISOFLIFD	620	634	BY	SIMILARITY
FT	DISOFLIFD	636	649	BY	SIMILARITY
FT	DISOFLIFD	655	666	BY	SIMILARITY
FT	DISOFLIFD	661	675	BY	SIMILARITY
FT	DISOFLIFD	667	680	BY	SIMILARITY
FT	DISOFLIFD	765	777	BY	SIMILARITY
FT	DISOFLIFD	772	786	BY	SIMILARITY
FT	DISOFLIFD	788	801	BY	SIMILARITY
FT	DISOFLIFD	807	819	BY	SIMILARITY
FT	DISOFLIFD	814	828	BY	SIMILARITY
FT	DISOFLIFD	830	843	BY	SIMILARITY
FT	DISOFLIFD	849	859	BY	SIMILARITY
FT	DISOFLIFD	854	868	BY	SIMILARITY
FT	DISOFLIFD	870	883	BY	SIMILARITY
FT	DISOFLIFD	952	964	BY	SIMILARITY
FT	DISOFLIFD	959	973	BY	SIMILARITY
FT	DISOFLIFD	975	988	BY	SIMILARITY
FT	DISOFLIFD	1070	1082	BY	SIMILARITY
FT	DISOFLIFD	1077	1091	BY	SIMILARITY
FT	DISOFLIFD	1093	1106	BY	SIMILARITY
FT	DISOFLIFD	1112	1124	BY	SIMILARITY
FT	DISOFLIFD	1119	1133	BY	SIMILARITY
FT	DISOFLIFD	1135	1149	BY	SIMILARITY
FT	DISOFLIFD	1155	1167	BY	SIMILARITY
FT	DISOFLIFD	1162	1176	BY	SIMILARITY
FT	DISOFLIFD	1178	1191	BY	SIMILARITY
FT	DISOFLIFD	1197	1209	BY	SIMILARITY
FT	DISOFLIFD	1204	1218	BY	SIMILARITY
FT	DISOFLIFD	1220	1233	BY	SIMILARITY
FT	DISOFLIFD	1239	1250	BY	SIMILARITY
FT	DISOFLIFD	1246	1259	BY	SIMILARITY
FT	DISOFLIFD	1261	1274	BY	SIMILARITY
FT	DISOFLIFD	1280	1292	BY	SIMILARITY
FT	DISOFLIFD	1287	1301	BY	SIMILARITY
FT	DISOFLIFD	1303	1316	BY	SIMILARITY
FT	DISOFLIFD	1322	1334	BY	SIMILARITY
FT	DISOFLIFD	1329	1343	BY	SIMILARITY
FT	DISOFLIFD	1345	1358	BY	SIMILARITY
FT	DISOFLIFD	1364	1377	BY	SIMILARITY
FT	DISOFLIFD	1371	1386	BY	SIMILARITY
FT	DISOFLIFD	1388	1399	BY	SIMILARITY
FT	DISOFLIFD	1405	1418	BY	SIMILARITY
FT	DISOFLIFD	1412	1427	BY	SIMILARITY
FT	DISOFLIFD	1429	1440	BY	SIMILARITY
FT	DISOFLIFD	1446	1458	BY	SIMILARITY
FT	DISOFLIFD	1453	1467	BY	SIMILARITY
FT	DISOFLIFD	1469	1482	BY	SIMILARITY
FT	DISOFLIFD	1484	1499	BY	SIMILARITY
FT	DISOFLIFD	1494	1508	BY	SIMILARITY
FT	DISOFLIFD	1510	1523	BY	SIMILARITY
FT	DISOFLIFD	1529	1540	BY	SIMILARITY
FT	DISOFLIFD	1535	1549	BY	SIMILARITY
FT	DISOFLIFD	1551	1564	BY	SIMILARITY
FT	DISOFLIFD	1647	1659	BY	SIMILARITY
FT	DISOFLIFD	1654	1668	BY	SIMILARITY
FT	DISOFLIFD	1670	1683	BY	SIMILARITY
FT	DISOFLIFD	1689	1701	BY	SIMILARITY
FT	DISOFLIFD	1696	1710	BY	SIMILARITY

Query Match 11.8%; Score 114.5; DB 1; Length 2907;  
 Best Local Similarity 23.7%; Pred. No. 0.0077;  
 Matches 47; Conservative 19; Mismatches 69; Indels 63; Gaps 11;

QY 1 MARHGRSLTTLVPMWAAALLALGERALALPEICT-OCPSGVNLSKVAFCYKTTREL 58  
 Db 2429 ICHHPRGYARD-----GRDIDCECKVMPSLCTNGOC---VNTMGSFRCCKVGYTM 2475  
 QY 59 MLARCCILNCKGTLIGLDLONCSLEDPGP-NFHQAHTT-----VIIDQANPLKG 107  
 Db 2476 DIGSTAC-----VDLDECS-OSPKPCNFTCKNTKSGYSCSPRGVLAQDCKTKC- 2524  
 QY 108 DL-----ANTFPGFTQLTLPLQHVNCPCGINAMNTITSYDNOICQCGOK 153  
 Db 2525 DLDECCQKQHNCCQFLCVNTLIGFT-----CKCPFGFTQHH--TACIDNNECGSQP 2572  
 QY 154 NLGNNGTDPCEMCEPENGSC 171  
 Db 2573 SLGAKGICONTPGSFSC 2590

RESULT 2  
 FBN2\_HUMAN STANDARD; PRT; 2911 AA.

AC P35356;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE FIBRILIN 2 PRECURSOR.  
 GN FBN2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94165150; PubMed=8120105;  
 RA Zhang H., Apfelroth S.D., Hu W., Davis E.C., Sanguineti C.,  
 RA Bondio J., Mecham R.P., Ramirez F.;  
 RT "Structure and expression of fibrillin-2, a novel microfibrillar  
 RT component preferentially located in elastic matrices.";  
 RT J. Cell Biol. 124:855-863(1994).  
 RL [2]  
 RN SEQUENCE OF 752-1505 FROM N.A.  
 RX MEDLINE=91304567; PubMed=1852206;  
 RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Safarazi M.,  
 RA Tsipouras P., Ramirez F., Hollister D.;  
 RT "Linkage of Marfan syndrome and a phenotypically related disorder to  
 RT two different fibrillin genes.";  
 RT Nature 352:330-334(1991).  
 RL [3]  
 RN VARIANTS CCA TYR-1252 AND SER-1433, AND VARIANT ILE-964.  
 RX MEDLINE=96083599; PubMed=7493032;  
 RA Putnam E.A., Zhang H., Ramirez F., Milewicz D.M.;  
 RT "Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder,  
 RT congenital structural arachnodactyly.";  
 RT Nat. Genet. 11:456-458(1995).  
 RL [4]  
 RN VARIANTS CCA HIS-1114.  
 RX MEDLINE=98407789; PubMed=9737771;  
 RA Babcock D., Gasner C., Francke U., Maslen C.;  
 RT "A single mutation that results in an asp-to-his substitution and  
 RT partial exon skipping in a family with congenital contractural  
 RT arachnodactyly.";  
 RT Hum. Genet. 103:22-28(1998).  
 RL [5]  
 RN VARIANTS CCA PHE-1141 AND TRP-1252.  
 RX MEDLINE=20259236; PubMed=10797416;  
 RA Belleh S., Zhou G., Wang M., Der Kaloustian V.M., Pagon R.A.,  
 RA Godfrey M.;  
 RT "Two novel fibrillin-2 mutations in congenital contractural  
 RT arachnodactyly.";

RL Am. J. Med. Genet. 92:7-12(2000).  
 CC -I- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS  
 CC THAT BINDS CALCIUM. FIBRILIN-2-CONTAINING MICROFIBRILS REGULATE  
 CC THE EARLY PROCESS OF ELASTIC FIBER ASSEMBLY.  
 CC -I- DISEASE: DEFECTS IN FBN2 ARE THE CAUSE OF CONGENITAL CONTRACTURAL  
 CC ARACHNOACTYLY (CCA) (ALSO KNOWN AS BEALS SYNDROME). CCA IS  
 CC PHENOTYPICALLY SIMILAR TO MARFAN SYNDROME, BUT DOES NOT AFFECT THE  
 CC AORTA AND THE EYES.  
 CC -I- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING  
 CC EGF-LIKE DOMAINS.  
 CC -I- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U03272; AAA18950.1; -;  
 DR EMBL: X62009; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: S17063; S17063.  
 DR PIR: S31101; S31101.  
 DR HSSP: P35555; IEMO.  
 DR MIM: 121050; -;  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR001438; EGF-II.  
 DR InterPro: IPR002212; TB.  
 DR Pfam: PF00008; EGF; 46.  
 DR Pfam: PF00683; TB; 9.  
 DR PRINTS: PR00010; EGFBLDOD.  
 DR SMART: SM00179; EGF-Ca; 43.  
 DR SMART: SM00001; EGF-like; 3.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 43.  
 DR PROSITE: PS00022; EGF\_1; 2.  
 DR PROSITE: PS01186; EGF\_2; 37.  
 DR PROSITE: PS01187; EGF-Ca; 43.  
 KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;  
 KW Repeat; Signal; Multigene family; Disease mutation; Polymorphism.  
 FT CHAIN 1  
 FT SIGNAL 1  
 FT 2911  
 FT 111  
 FT DOMAIN 1  
 FT 145  
 FT DOMAIN 1  
 FT 176  
 FT DOMAIN 1  
 FT 275  
 FT DOMAIN 1  
 FT 316  
 FT DOMAIN 1  
 FT 358  
 FT REPEAT 1  
 FT 425  
 FT DOMAIN 1  
 FT 493  
 FT 534  
 FT DOMAIN 1  
 FT 574  
 FT DOMAIN 1  
 FT 615  
 FT DOMAIN 1  
 FT 656  
 FT REPEAT 1  
 FT 697  
 FT DOMAIN 1  
 FT 767  
 FT 808  
 FT DOMAIN 1  
 FT 850  
 FT 890  
 FT DOMAIN 1  
 FT 954  
 FT 995  
 FT REPEAT 1  
 FT 1071  
 FT 1113  
 FT DOMAIN 1  
 FT 1156  
 FT DOMAIN 1  
 FT 1198  
 FT DOMAIN 1  
 FT 1241  
 FT 1281  
 FT DOMAIN 1  
 FT 1323  
 FT DOMAIN 1  
 FT 1365  
 FT DOMAIN 1  
 FT 1406  
 FT 1447  
 FT DOMAIN 1  
 FT 1489  
 FT 1530  
 FT DOMAIN 1  
 FT 1571

```

FT REPEAT 1572 1648 TGRBP 4.
FT DOMAIN 1649 1690 EGF-LIKE 27, CALCIUM-BINDING.
FT REPEAT 1691 1732 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1733 1806 TGRBP 5.
FT DOMAIN 1807 1848 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1891 1932 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1933 1971 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1972 2014 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 2015 2054 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2055 2096 EGF-LIKE 35, CALCIUM-BINDING.
FT REPEAT 2097 2169 TGRBP 6.
FT DOMAIN 2170 2211 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2212 2251 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2252 2292 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2293 2336 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2337 2378 EGF-LIKE 40, CALCIUM-BINDING.
FT REPEAT 2379 2447 TGRBP 7.
FT DOMAIN 2448 2489 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2490 2530 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2531 2569 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2570 2612 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2613 2652 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2653 2693 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2694 2733 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 115 124 BY SIMILARITY.
FT DISULFID 119 130 BY SIMILARITY.
FT DISULFID 132 141 BY SIMILARITY.
FT DISULFID 149 159 BY SIMILARITY.
FT DISULFID 153 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 180 190 BY SIMILARITY.
FT DISULFID 184 195 BY SIMILARITY.
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 279 291 BY SIMILARITY.
FT DISULFID 286 300 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 321 333 BY SIMILARITY.
FT DISULFID 328 342 BY SIMILARITY.
FT DISULFID 344 357 BY SIMILARITY.
FT DISULFID 497 509 BY SIMILARITY.
FT DISULFID 504 518 BY SIMILARITY.
FT DISULFID 520 532 BY SIMILARITY.
FT DISULFID 538 548 BY SIMILARITY.
FT DISULFID 543 557 BY SIMILARITY.
FT DISULFID 559 572 BY SIMILARITY.
FT DISULFID 578 590 BY SIMILARITY.
FT DISULFID 585 599 BY SIMILARITY.
FT DISULFID 601 614 BY SIMILARITY.
FT DISULFID 620 631 BY SIMILARITY.
FT DISULFID 626 640 BY SIMILARITY.
FT DISULFID 642 655 BY SIMILARITY.
FT DISULFID 661 672 BY SIMILARITY.
FT DISULFID 667 681 BY SIMILARITY.
FT DISULFID 683 696 BY SIMILARITY.
FT DISULFID 771 783 BY SIMILARITY.
FT DISULFID 778 792 BY SIMILARITY.
FT DISULFID 794 807 BY SIMILARITY.
FT DISULFID 813 825 BY SIMILARITY.
FT DISULFID 820 834 BY SIMILARITY.
FT DISULFID 836 849 BY SIMILARITY.
FT DISULFID 855 865 BY SIMILARITY.
FT DISULFID 860 874 BY SIMILARITY.
FT DISULFID 876 889 BY SIMILARITY.
FT DISULFID 970 979 BY SIMILARITY.
FT DISULFID 958 979 BY SIMILARITY.
FT DISULFID 965 979 BY SIMILARITY.
FT DISULFID 981 994 BY SIMILARITY.
FT DISULFID 1076 1088 BY SIMILARITY.
FT DISULFID 1083 1097 BY SIMILARITY.
FT DISULFID 1099 1112 BY SIMILARITY.
FT DISULFID 1118 1130 BY SIMILARITY.
FT DISULFID 1125 1139 BY SIMILARITY.
FT DISULFID 1141 1155 BY SIMILARITY.

```

```

FT DISULFID 1161 1173 BY SIMILARITY.
FT DISULFID 1168 1182 BY SIMILARITY.
FT DISULFID 1184 1197 BY SIMILARITY.
FT DISULFID 1203 1215 BY SIMILARITY.
FT DISULFID 1210 1224 BY SIMILARITY.

Query Match 10.9% Score 105.5; DB 1; Length 2911;
Best local Similarity 22.7%; Pred. No. 0.057; 71; Indels 63; Gaps 11;
Matches 45; Conservative 19; Mismatches 71; Indels 63; Gaps 11;

QY 1 MAPHGSGTITTPVMAAALLALGVERALALPEICT--QCPGSSYONLSKVAFCYKTRREL 58
Db 2435 ICPHGSGYTTD-----GRDIDCKVMPNLCITNGCC---INTMSFRCFCVGYTT 2481
QY 59 MHAHRCCLNOKGTILGLDONSLEDPGF-NHQAHIT-----VIIDLANPIKLG 107
Db 2482 DISGTSCT-----IDDECS-OSPKCNVICKNTBGSYOCSPRGVLOEDGKTK- 2530
QY 108 DL-----ANFRGFTOLQTLILPQVNCPGINAMWNTSYIDNOICOGOK 153
Db 2531 DLDECGTKQHNCOFLCVNLTGFT-----CKCPGFTQHH--TACIDNNCGSGQP 2578
QY 154 NLGNNTGDEPMCPENGSC 171
Db 2579 LLGGGKIGICQNTPGSFSC 2596

RESULT 3
FBN1_HUMAN STANDARD; PRT; 2871 AA.
ID FBN1_HUMAN
AC P35555;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FIBRILLIN 1 PRECURSOR.
GN FBN1 OR FBN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93372860; PubMed=8364578;
RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA Pangillan T., Bonadio J.;
RT "Genomic organization of the sequence coding for fibrillin, the
RT defective gene product in Marfan syndrome."
RL Hum. Mol. Genet. 2:961-968(1993).
RN [2]
RP SEQUENCE OF 1-932 FROM N.A.
RC TISSUE=Placenta, and Fibroblast;
RX MEDLINE=94010947; PubMed=7691719;
RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;
RT "Fibrillin binds calcium and is coded by cDNAs that reveal a
RT multidomain structure and alternatively spliced exons at the 5'
RT end."
RL Nature 352:334-337(1991).
RN [3]
RP SEQUENCE OF 899-2871 FROM N.A.
RX MEDLINE=91304568; PubMed=1852207;
RA Masten C.L., Corson G.M., Maddox B.K., Gianville R.W., Sakai L.Y.;
RT "Partial sequence of a candidate gene for the Marfan syndrome."
RL Nature 352:334-337(1991).
RN [4]
RP SEQUENCE OF 813-1313 FROM N.A.
RX MEDLINE=91304567; PubMed=1852206;
RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.G., Sarfarazi M.,
RA Tsipouras P., Ramirez F., Hollister D.W.;
RT "Linkage of Marfan syndrome and a phenotypically related disorder to
RT two different fibrillin genes."
RL Nature 352:330-334(1991).
RN [5]

```

RP CHARACTERIZATION. PubMed=1860873;  
 RX MEDLINE=91317849; Glanville R.W., Baehlinger H.P.;  
 RA Sakai L.Y., Keene D.R., Glanville R.W., Baehlinger H.P.;  
 RT "Purification and partial characterization of fibrillin, a cysteine-  
 RT rich structural component of connective tissue microfibrils.";  
 RL J. Biol. Chem. 266:14763-14770(1991).  
 RN [6]  
 RP STRUCTURE BY NMR OF 2054-2125.  
 RX MEDLINE=98031893; PubMed=9362480;  
 RA Yuan X., Downing A.K., Knott V., Handford P.A.;  
 RT "Solution structure of the transforming growth factor beta-binding  
 RT protein-like module, a domain associated with matrix fibrils.";  
 RL EMBO J. 16:6659-6666(1997).  
 RN [7]  
 RP STRUCTURE BY NMR OF 2124-2205.  
 RX MEDLINE=96144829; PubMed=8568869;  
 RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;  
 RT "Calcium binding properties of an epidermal growth factor-like domain  
 RT pair from human fibrillin-1.";  
 RL J. Mol. Biol. 255:22-27(1996).  
 RN [8]  
 RP STRUCTURE BY NMR OF 2124-2205.  
 RX MEDLINE=96222301; PubMed=8653794;  
 RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,  
 RA Handford P.A.;  
 RT "Solution structure of a pair of calcium-binding epidermal growth  
 RT factor-like domains: implications for the Marfan syndrome and other  
 RT genetic disorders.";  
 RL Cell 85:597-605(1996).  
 RN [9]  
 RP REVIEW ON MFS VARIANTS.  
 RX MEDLINE=96174615; PubMed=8594563;  
 RA Collod G., Beroud C., Soussi T., Junien C., Boileau C.;  
 RT "Software and database for the analysis of mutations in the human  
 RT FBN1 gene.";  
 RL Nucleic Acids Res. 24:137-141(1996).  
 RN [10]  
 RP REVIEW ON MFS VARIANTS.  
 RX MEDLINE=97169383; PubMed=9015626;  
 RA Collod-Beroud G., Beroud C., Ades L., Black C., Boxer M., Brock D.J.,  
 RA Godfrey M., Hayward C., Karttunen L., Millewicz D., Peltonen L.,  
 RA Richards R.I., Wang W., Junien C., Boileau C.;  
 RT "Marfan Database (second edition): software and database for the  
 RT analysis of mutations in the human FBN1 gene.";  
 RL Nucleic Acids Res. 25:147-150(1997).  
 RN [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=98062175; PubMed=9401003;  
 RA Hayward C., Brock D.J.H.;  
 RT "Fibrillin-1 mutations in Marfan syndrome and other type-1  
 RT fibrilopathies.";  
 RL Hum. Mutat. 10:415-423(1997).  
 RN [12]  
 RP VARIANT MFS PRO-1137.  
 RX MEDLINE=91304569; PubMed=1855208;  
 RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,  
 RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,  
 RA Curstain S.M., Stetten G., Meyers D.A., Francomano C.A.;  
 RT "Marfan syndrome caused by a recurrent de novo missense mutation in  
 RT the fibrillin gene.";  
 RL Nature 352:337-339(1991).  
 RN [13]  
 RP VARIANTS MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.  
 RX MEDLINE=93250834; PubMed=1301946;  
 RA Dietz H.C., Saraiva J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;  
 RT "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome  
 RT patients at cysteine residues in EGF-like domains.";  
 RL Hum. Mutat. 1:366-374(1992).  
 RN [14]  
 RP VARIANT MFS SER-2307.  
 RX MEDLINE=92235290; PubMed=1569206;  
 RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendler R.J., Jr.,  
 RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;  
 RT "Marfan phenotype variability in a family segregating a missense  
 RT mutation in the epidermal growth factor-like motif of the fibrillin  
 RT gene.";  
 RL J. Clin. Invest. 89:1674-1680(1992).  
 RN [15]  
 RP VARIANTS MFS ILE-548 AND ALA-723.  
 RX MEDLINE=94010946; PubMed=8406497;  
 RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,  
 RA Pyeritz R.E., Francomano C.A.;  
 RT "Four novel FBN1 mutations: significance for mutant transcript level  
 RT and EGF-like domain calcium binding in the pathogenesis of Marfan  
 RT syndrome.";  
 RL Genomics 17:468-475(1993).  
 RN [16]  
 RP VARIANTS MFS SER-2144.  
 RX MEDLINE=93278402; PubMed=8504310;  
 RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;  
 RT "A novel fibrillin mutation in the Marfan syndrome which could  
 RT disrupt calcium binding of the epidermal growth factor-like module.";  
 RL Hum. Mol. Genet. 2:475-477(1993).  
 RN [17]  
 RP VARIANTS MFS R-862; Y-1117; P-1137 AND F-1589, AND VARIANT A-1148.  
 RX MEDLINE=94108431; PubMed=8281141;  
 RA Tyan K., Comeau K., Pearson M., Willgenbus P., Levitt D., Gasner C.,  
 RA Berg M.A., Miller D.C., Francke U.;  
 RT "Mutation screening of complete fibrillin-1 coding sequence: report  
 RT of five new mutations, including two in 8-cysteine domains.";  
 RL Hum. Mol. Genet. 2:1813-1821(1993).  
 RN [18]  
 RP VARIANTS MFS GLY-217 AND ARG-2627.  
 RX MEDLINE=95067970; PubMed=7973366;  
 RA Karttunen L., Raghunath M., Loeonqvist L., Peltonen L.;  
 RT "A compound heterozygous Marfan patient: two defective fibrillin  
 RT alleles result in a lethal phenotype.";  
 RL Am. J. Hum. Genet. 55:1083-1091(1994).  
 RN [19]  
 RP VARIANT EL LYS-2447.  
 RX MEDLINE=94245249; PubMed=8188302;  
 RA Loeonqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,  
 RA Peltonen L.;  
 RT "A novel mutation of the fibrillin gene causing ectopia lentis.";  
 RL Genomics 19:573-576(1994).  
 RN [20]  
 RP VARIANT MFS CYS-627.  
 RX MEDLINE=94272487; PubMed=8004112;  
 RA Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;  
 RT "Two novel mutations and a neutral polymorphism in EGF-like domains  
 RT of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan  
 RT syndrome patients.";  
 RL Hum. Mol. Genet. 3:373-375(1994).  
 RN [21]  
 RP VARIANT MFS CYS-122.  
 RX MEDLINE=94314977; PubMed=8040326;  
 RA Stahl-Hallenqren C., Ukonen T., Kainulainen K., Kristofersson U.,  
 RA Saxne T., Torngvist K., Peltonen L.;  
 RT "An extra cysteine in one of the non-calcium-binding epidermal growth  
 RT factor-like motifs of the FBN1 polypeptide is connected to a novel  
 RT variant of Marfan syndrome.";  
 RL J. Clin. Invest. 94:709-713(1994).  
 RN [22]  
 RP VARIANT MFS TYR-1223.  
 RX MEDLINE=94351682; PubMed=8071963;  
 RA Hewett D.R., Lynch J.R., Child A., Sykes B.C.;  
 RT "A new missense mutation of fibrillin in a patient with Marfan  
 RT syndrome.";  
 RL J. Med. Genet. 31:338-339(1994).  
 RN [23]  
 RP VARIANT MFS HIS-1170.  
 RX MEDLINE=95174777; PubMed=7870075;  
 RA Hayward C., Porteous M.E.M., Brock D.J.H.;  
 RT "A novel mutation in the fibrillin gene (FBN1) in familial  
 RT atachnodyly.";  
 RL Mol. Cell. Probes 8:325-327(1994).

RN [24]  
 RP VARIANTS MES G217:N1023;R1074;Y1242;R1513;E2127;W2151;K2447 AND R2511  
 RX MEDLINE=94164368; PubMed=8136837;  
 RA Kainulainen K., Karttunen L., Puhaka L., Sakai L., Peltonen L.;  
 RT "Mutations in the fibrillin gene responsible for dominant ectopia  
 RL lentis and neonatal Marfan syndrome."; Nat. Genet. 6:64-69(1994).  
 RN [25]  
 RP VARIANT SER-1127.  
 RX MEDLINE=95282774; PubMed=7762551;

Query Match	9.8%	Score 95.5	DB 1	Length 2871
Best Local Similarity	21.4%	Pred. No. 0.53		
Matches	43	Conservative	20	Mismatches 69; Indels 69; Gaps 9

QY	1	MARPGSLTTLVPM	MAALLALG	VERALALPEICT	-QC	----	PGSVQNLISVA	-----	49
				:	:	:			
Db	2389	LCPHRGFMTN	-----	GADIDECKV	HDVCRN	CECVNDR	SGSYHC	ICKTGYPTDT	2438
QY	50	-----	----	FYCKTREL	MLHARC	CINOKETI	IGLDI	ONC-SLED	QGNPH 90
Db	2439	GTSQVDLNEC	NOAPR	PCNF	ICKNTE	-----	GSYGQSC	PKRYII	QEDORSK
									DIKTRKH 2494
QY	91	QAHTTVIIDL	QANPLK	GLDANT	FRGFT	QIQTLL	PDHVN	CPGGIN	AMNTITSY
									IDNQIQ 150
Db	2495	NCQFLCV	-----	NTIGFT	-----	CKSPGFT	QOHN	-TSC	IDINNECT
									2529
QY	151	GQKNLCNN	TGDEPM	CENG	SC 171				
Db	2530	SPDNLG	SGKIGI	CQNT	PGSFTC	2550			

RESULT	4			
TCF8_CHICK		STANDARD;		PRT; 1114 AA.
ID	TCF8_CHICK			
AC	P36197; 042408;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	TRANSCRIPTION FACTOR 8 (DELTA-CRYSTALLIN ENHANCER BINDING FACTOR) (DELTA EBF).			
OS	Gallus gallus (chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lens;			
RA	MEDLINE=94116444; Pubmed=7904558;			
RA	Funahashi J.-I., Sekido R., Murai K., Kamachi Y., Kondoh H.;			
RT	"Delta-crystallin enhancer binding protein delta EBF is a zinc			
RT	finger-homeodomain protein implicated in postgastrulation			
RT	embryogenesis.";			
RL	Development 119:433-446(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Embryo;			
RA	MEDLINE=97082972; Pubmed=8964504;			
RA	Sekido R., Takagi T., Okanami M., Moribe H., Yamamura M., Higashi Y.,			
RA	Kondoh H.;			
RT	"Organization of the gene encoding transcriptional repressor deltaEBF1			
RT	and cross-species conservation of its domains.";			
RL	Gene 173:227-232(1996).			
CC	-1- FUNCTION: BINDS TO DELTA 1-CRYSTALLIN ENHANCER CORE AND REPRESSED			
CC	LENS-SPECIFIC TRANSCRIPTION. IT BINDS AS WELL MANY OTHER NON-LENS			
CC	SPECIFIC DNA SEQUENCES.			
CC	-1- SUBUNIT: MAY INTERACT WITH CTBP1 (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).			
CC	-1- TISSUE SPECIFICITY: EXPRESSION IS DEVELOPMENTALLY REGULATED WITH			
CC	HIGH EXPRESSION IN MESODERM, NERVOUS SYSTEM AND LENS.			
CC	-1- DEVELOPMENTAL STAGE: EXPRESSION STARTS AFTER GASTRULATION, WHEN			
CC	ORANGENESIS HAS JUST BEGUN.			

CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZNF-1 FAMILY OF TWO-HANDED ZINC  
CC FINGER/HOMEDODOMAIN PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL; D14313; BAA03259.1;	-
DR	EMBL; D76434; BAAL1178.1;	-
DR	EMBL; D76433; BAAL1178.1;	JOINED.
DR	HSSP; P08047; 1SP2.	
DR	TRANSECAC; T01467;	-
DR	InterPro; IPR001356;	Homeobox.
DR	InterPro; IPR000822;	Znf-C2H2.
DR	Pfam; PF00096;	Zf-C2H2; 7.
DR	PRINTS; PR00048;	ZINCFLINGER.
DR	SMART; SM00389;	HOX.1.
DR	SMART; SM00355;	Znf-C2H2; 7.
DR	PROSITE; PS00028;	ZINC_FINGER_C2H2_1; 5.
DR	PROSITE; PS50157;	ZINC_FINGER_C2H2_2; 7.
KW	Transcription regulation; DNA-binding; Nuclear protein; zinc-finger; Homeobox; Metal-binding; Repeat.	
KW	ZN_FING	170 193 C2H2-TYPE.
FT	ZN_FING	200 222 C2H2-TYPE.
FT	ZN_FING	240 262 C2H2-TYPE.
FT	ZN_FING	268 292 C2HC-TYPE (ATYPICAL).
FT	DNA_BIND	581 640 HOMEBOX-LIKE.
FT	SITE	767 771 CTBP-BINDING MOTIF (BY SIMILARITY).
FT	ZN_FING	904 926 C2H2-TYPE.
FT	ZN_FING	932 954 C2H2-TYPE.
FT	ZN_FING	960 981 C2H2-TYPE (ATYPICAL).
FT	DOMAIN	989 1081 GLU-RICH (ACIDIC).
FT	CONFLICT	299 299 K -> N (IN REF. 2).
SEQ	SEQUENCE	1114 AA; 123150 MW; D914284143E7D279 CRC64;

	Query Match	9.4%;	Score 91;	DB 1;	Length 114;	
	Best Local Similarity	21.3%;	Pred. No. 0.5;			
Matches	42;	Conservative	22;	Mismatches	73;	Gaps
					Indels	60;
						Gaps
OY	24 GERALALPEITOCGP-----GSVONLSKVAFYCKTTRRLM--LHARCLNMQGTIIGLD	76				
	: :   :   :   :   :   :   :   :   :   :   :					
Db	389 GVQAQAVLPETVGVSPTSINLSIDIGNLVKAVDGNVIRQVLNNHANLASKEOETISNAS	448				
OY	77 LQ-----NCSLEDPG-----PNFQAHTTYIIDDQANPL	105				
	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :					
Db	449 IQAQASHLSAISLPVDDGGTKIIITIVSLSPSQLOLVFONLKREHSVPPTNSCKNEKL	508				
OY	106 KGDLA-----NTFRGETQLQTLIPHOHVCPGGIANMTITSYIDNOICOGOKNLICNNT	159				
	:					
Db	509 PEDLVAKSEKDKNFGEFTNDSTCLLCD--DCPGDINALQELKHV-----ET	552				
OY	160 GDPENCPCNGSCVPPDGP	176				
	:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :					
Db	553 KNPPOLPOSGTEAEKP	569				
RESULT	5					
Y053_HUMAN						
ID	Y053_HUMAN	STANDARD:	PRT:	638 AA.		
AC	P42331.					
DT	01-NOV-1995 (rel. 32, Created)					
DT	01-NOV-1995 (rel. 32, Last sequence update)					
DT	20-AUG-2001 (rel. 40, Last annotation update)					
DE	HYPOTHETICAL PROTEIN KIAA0053.					
GN	KIAA0053.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.					



```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. 11.
RT The coding sequences of 40 new genes (K1A0041-K1A0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: D29642; BAA06125.1; -
CC HSSP: 007960; IRGP.
CC InterPro: IPR001849; PH.
CC InterPro: IPR00198; RhogAP.
CC Pfam: PF00169; PH; 1.
CC SMART: SM00233; RhogAP; 1.
CC SMART: SM00324; PH; 1.
CC PROSITE: PS00003; PH_DOMAIN; 1.
CC DR Hypothetical protein.
CC KM
CC FT DOMAIN 39 144 PH.
CC S0 SEQUENCE 638 AA; 72431 MW; 0B2897FBA4880871 CRC64;

```

```

Query Match          9.1%; Score 88.5; DB 1; Length 638;
Best Local Similarity 21.7%; Pred. NO. 0.47;
Matches 44; Conservative 38; Mismatches 66; Indels 55; Gaps 11;

```

```

OY 12 LVFWA--AALLLA--LGERALALPEICTQCGSV--QNSKVAFCCKTRELMLHAR 63
DB 239 VVPMQSYEGFLLCGQLTNDEAKAOELMKQL--SILPRDNTSLSTYCRFLHEIQLN-- 294
OY 64 CCLNOK-----GTIIGLDLQNSLEDPG-----PNEHQATTVIILQA----- 102
DB 295 CAVNKMVSVDNLATVIGVNLIRSKVEDPAVMKGPQIQVMTMIMROHEVLFPSKIDPL 354
OY 103 -----NFLKGDLANTFPGFTQLOTLILPQVNVCPGGINAMNTTSTIDNOCGOKNL 155
DB 355 SEPAAKNNDKKAPVARSSVGMDATEDLRISR-----TDSFSSKMTSDSPTSPGOO-- 405
OY 156 CNNTGDEPCPENGSCVP--DGPG 177
DB 406 -----PSDAFPEPSSSKVPREKPG 423

```

```

RESULT 6
FBN1_MOUSE
ID FBN1_MOUSE STANDARD; PRT: 2871 AA.
AC Q61554; Q60826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FIBRILLIN 1 PRECURSOR.
GN FBN1 OR FBN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130561; PubMed=7829516;
RX Yin W., Germiller J., Sanguinetti C., Smiley E., Panglisan T.,

```

```

RA Pereira L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene.";
RL J. Biol. Chem. 270:1798-1806(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RA Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT
CC -1- PTM: FORMS INTRAMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: L29454; AAA56840.1; -
CC EMBL: U22493; AAA64217.1; -
CC HSSP: P35555; IABP.
CC MOD: MG1:95489; Fbn1.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001881; EGF-Ca.
CC InterPro: IPR002212; TB.
CC Pfam: PF00000; EGF; 46.
CC SMART: SM00000; TB; 9.
CC SMART: SM00000; EGF-Ca; 42.
CC PROSITE: PS00010; ASX_HYDROXYL; 43.
CC PROSITE: PS00022; EGF_1; 2.
CC PROSITE: PS01186; EGF_2; 38.
CC PROSITE: PS01187; EGF-Ca; 45.
CC Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
CC Repeat; Signal; Multigene family.
CC STANAL.
FT CHAIN 1 2871
FT DOMAIN 81 112 EGF-LIKE 1, NON-CALCIUM BINDING.
FT DOMAIN 115 146 EGF-LIKE 2, NON-CALCIUM BINDING.
FT DOMAIN 147 178 EGF-LIKE 3, NON-CALCIUM BINDING.
FT DOMAIN 188 287 EGF-LIKE 4, CALCIUM BINDING.
FT DOMAIN 288 329 EGF-LIKE 5, CALCIUM BINDING.
FT DOMAIN 330 401 TGFBP 1.
FT DOMAIN 402 446 PRO-RICH.
FT DOMAIN 447 489 EGF-LIKE 6, NON-CALCIUM BINDING.
FT DOMAIN 490 529 EGF-LIKE 7, CALCIUM BINDING.
FT DOMAIN 530 571 EGF-LIKE 8, CALCIUM BINDING.
FT DOMAIN 572 612 EGF-LIKE 9, CALCIUM BINDING.
FT DOMAIN 613 653 EGF-LIKE 10, CALCIUM BINDING.
FT DOMAIN 654 721 TGFBP 2.
FT DOMAIN 722 764 EGF-LIKE 11, CALCIUM BINDING.
FT DOMAIN 765 806 EGF-LIKE 12, CALCIUM BINDING.
FT DOMAIN 807 846 EGF-LIKE 13, CALCIUM BINDING.
FT DOMAIN 847 910 EGF-LIKE 14, CALCIUM BINDING.
FT DOMAIN 911 951 TGFBP 3.
FT DOMAIN 952 1018 EGF-LIKE 15, CALCIUM BINDING.
FT DOMAIN 1019 1069 EGF-LIKE 16, CALCIUM BINDING.
FT DOMAIN 1070 1112 EGF-LIKE 17, CALCIUM BINDING.
FT DOMAIN 1113 1154 EGF-LIKE 18, CALCIUM BINDING.
FT DOMAIN 1155 1196 EGF-LIKE 19, CALCIUM BINDING.
FT DOMAIN 1197 1237 EGF-LIKE 20, CALCIUM BINDING.
FT DOMAIN 1238 1279 EGF-LIKE 21, CALCIUM BINDING.
FT DOMAIN 1280 1321 EGF-LIKE 22, CALCIUM BINDING.
FT DOMAIN 1322 1362

```

```

FT DOMAIN 1363 1403 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1404 1445 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1446 1486 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1487 1527 EGF-LIKE 26, CALCIUM-BINDING.
FT DOMAIN 1528 1599 TGFBRP 4.
FT DOMAIN 1608 1647 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1648 1688 EGF-LIKE 28, CALCIUM-BINDING.
FT DOMAIN 1689 1758 TGFBRP 5.
FT DOMAIN 1766 1807 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1808 1848 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1891 1929 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1930 1972 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 1973 2012 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2013 2054 EGF-LIKE 35, CALCIUM-BINDING.
FT DOMAIN 2055 2121 TGFBRP 6.
FT DOMAIN 2127 2165 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2166 2205 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2206 2246 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2247 2290 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2291 2332 EGF-LIKE 40, CALCIUM-BINDING.
FT DOMAIN 2333 2400 TGFBRP 7.
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2485 2523 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2567 2606 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2648 2687 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 89 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
FT DISULFID 119 129 BY SIMILARITY.
FT DISULFID 123 134 BY SIMILARITY.
FT DISULFID 136 145 BY SIMILARITY.
FT DISULFID 150 160 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 250 262 BY SIMILARITY.
FT DISULFID 271 286 BY SIMILARITY.
FT DISULFID 292 304 BY SIMILARITY.
FT DISULFID 299 313 BY SIMILARITY.
FT DISULFID 315 328 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 474 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 776 790 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 830 845 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.

```

```

FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.

```

## Query Match

9.1%; Score 88.5; DB 1; Length 2871;

Best Local Similarity 20.9%; Pred. No. 2.5; Matches 42; Conservative 20; Mismatches 70; Indels 69; Gaps 9;

```

QY 1 MAPHGSLTTLVPMWAAALLALGVERALALPETCT--OC---PGSVQMLSKVA----- 49
DB 2389 LCPHGRGPMNT-----GADVDECKYIHVCRNGECVNDRGSHCICKTGYTPDIT 2438
QY 50 -----FYCKTTRRLMLHARCCLNMGKFTILGIDLONC-SLEDGPMPFH 90
DB 2439 GTSVCVDINECNQAPKPCNFICKNTE---GSYQSCSPNGYITLQEDGRCKDDECATKQH 2494
QY 91 QAHFTVITIDQANPLKGDLANFERGFTOTLTLLPQHVNCPGSINAMNTTSYIDNOICQ 150
DB 2495 NCGFLCV-----NTIGGFT-----CCKPFGFTQHH--TACIDNNECT 2529
QY 151 GOKNLGNNTGDEPCMPENGSC 171
DB 2530 SDINLCSGSKGICQNTPSFTC 2550

```

## RESULT 7

FBN1\_BOVIN STANDARD; PRT; 2871 AA.

```

AC P98133;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FIBRILIN 1 PRECURSOR (M9340).
GN FBN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;

```

FT	REPEAT	952	1027	TEGBP 3.	
FT	DOMAIN	1028	1069	EGF-LIKE 15,	CALCIUM-BINDING
FT	DOMAIN	1070	1112	EGF-LIKE 16,	CALCIUM-BINDING
FT	DOMAIN	1113	1154	EGF-LIKE 17,	CALCIUM-BINDING
FT	DOMAIN	1155	1196	EGF-LIKE 18,	CALCIUM-BINDING
FT	DOMAIN	1197	1237	EGF-LIKE 19,	CALCIUM-BINDING
FT	DOMAIN	1238	1279	EGF-LIKE 20,	CALCIUM-BINDING
FT	DOMAIN	1280	1321	EGF-LIKE 21,	CALCIUM-BINDING
FT	DOMAIN	1322	1362	EGF-LIKE 22,	CALCIUM-BINDING
FT	DOMAIN	1363	1403	EGF-LIKE 23,	CALCIUM-BINDING
FT	DOMAIN	1404	1445	EGF-LIKE 24,	CALCIUM-BINDING
FT	DOMAIN	1446	1486	EGF-LIKE 25,	CALCIUM-BINDING
FT	DOMAIN	1487	1527	EGF-LIKE 26,	CALCIUM-BINDING
FT	REPEAT	1528	1605	TEGBP 4.	
FT	DOMAIN	1606	1647	EGF-LIKE 27,	CALCIUM-BINDING
FT	DOMAIN	1648	1688	EGF-LIKE 28,	CALCIUM-BINDING
FT	REPEAT	1689	1765	TEGBP 5.	
FT	DOMAIN	1766	1807	EGF-LIKE 29,	CALCIUM-BINDING
FT	DOMAIN	1808	1848	EGF-LIKE 30,	CALCIUM-BINDING
FT	DOMAIN	1849	1890	EGF-LIKE 31,	CALCIUM-BINDING
FT	DOMAIN	1891	1929	EGF-LIKE 32,	CALCIUM-BINDING
FT	DOMAIN	1930	1972	EGF-LIKE 33,	CALCIUM-BINDING
FT	DOMAIN	1973	2012	EGF-LIKE 34,	CALCIUM-BINDING
FT	DOMAIN	2013	2054	EGF-LIKE 35,	CALCIUM-BINDING
FT	REPEAT	2055	2126	TEGBP 6.	
FT	DOMAIN	2127	2165	EGF-LIKE 36,	CALCIUM-BINDING
FT	DOMAIN	2166	2205	EGF-LIKE 37,	CALCIUM-BINDING
FT	DOMAIN	2206	2246	EGF-LIKE 38,	CALCIUM-BINDING
FT	DOMAIN	2247	2290	EGF-LIKE 39,	CALCIUM-BINDING
FT	DOMAIN	2291	2332	EGF-LIKE 40,	CALCIUM-BINDING
FT	REPEAT	2333	2401	TEGBP 7.	
FT	DOMAIN	2402	2443	EGF-LIKE 41,	CALCIUM-BINDING
FT	DOMAIN	2444	2484	EGF-LIKE 42,	CALCIUM-BINDING
FT	DOMAIN	2485	2523	EGF-LIKE 43,	CALCIUM-BINDING
FT	DOMAIN	2524	2566	EGF-LIKE 44,	CALCIUM-BINDING
FT	DOMAIN	2567	2606	EGF-LIKE 45,	CALCIUM-BINDING
FT	DOMAIN	2607	2647	EGF-LIKE 46,	CALCIUM-BINDING
FT	DOMAIN	2648	2687	EGF-LIKE 47,	CALCIUM-BINDING
FT	DISULFID	85	94	BY SIMILARITY.	
FT	DISULFID	89	100	BY SIMILARITY.	
FT	DISULFID	102	111	BY SIMILARITY.	
FT	DISULFID	119	129	BY SIMILARITY.	
FT	DISULFID	123	134	BY SIMILARITY.	
FT	DISULFID	136	145	BY SIMILARITY.	
FT	DISULFID	150	160	BY SIMILARITY.	
FT	DISULFID	154	166	BY SIMILARITY.	
FT	DISULFID	168	177	BY SIMILARITY.	
FT	DISULFID	250	262	BY SIMILARITY.	
FT	DISULFID	257	271	BY SIMILARITY.	
FT	DISULFID	273	286	BY SIMILARITY.	
FT	DISULFID	292	304	BY SIMILARITY.	
FT	DISULFID	313	325	BY SIMILARITY.	
FT	DISULFID	15	328	BY SIMILARITY.	
FT	DISULFID	453	465	BY SIMILARITY.	
FT	DISULFID	460	474	BY SIMILARITY.	
FT	DISULFID	476	488	BY SIMILARITY.	
FT	DISULFID	494	504	BY SIMILARITY.	
FT	DISULFID	499	513	BY SIMILARITY.	
FT	DISULFID	515	528	BY SIMILARITY.	
FT	DISULFID	534	545	BY SIMILARITY.	
FT	DISULFID	541	555	BY SIMILARITY.	
FT	DISULFID	557	570	BY SIMILARITY.	
FT	DISULFID	576	587	BY SIMILARITY.	
FT	DISULFID	582	596	BY SIMILARITY.	
FT	DISULFID	598	611	BY SIMILARITY.	
FT	DISULFID	617	628	BY SIMILARITY.	
FT	DISULFID	623	637	BY SIMILARITY.	
FT	DISULFID	639	652	BY SIMILARITY.	
FT	DISULFID	727	739	BY SIMILARITY.	
FT	DISULFID	734	748	BY SIMILARITY.	
FT	DISULFID	750	763	BY SIMILARITY.	
FT	DISULFID	769	781	BY SIMILARITY.	
FT	DISULFID	776	790	BY SIMILARITY.	

```

FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1245 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1336 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.

```

Query Match 8.9%; Score 86.5; DB 1; Length 2871;  
 Best Local Similarity 20.4%; Pred. No. 3.9; 70; Indels 69; Gaps 9;  
 Matches 41; Conservative 21; Mismatches

```

QY 1 MAPHGPSTLTTPWMAALLAGVERALALPEICT--QC--PGSVONLSKVA----- 49
DB 2389 LCPHGRGEMTN-----GADIDCEKVIHYDVCNNGCVDNRSGYHICKTGTPTDIT 2438
QY 50 -----FYCKTTRRLMHARCLLNQKFTTIGDLQNC-SLEDPGPNFH 90
DB 2439 GTACVLDNECNOAPKPCNFICKNTE-----GSYQSCPCPKYIIQDEGRSKDLDECATKOH 2494
QY 91 QAHRTTVIIDQANPLAKGDLNTRFGTQTLTLPVHVCNPGINAMNNTSYINQJCO 150
DB 2495 NCOFLCV-----NTGSEFT-----CKCPREFTQHH--TACIDNNECT 2529
QY 151 GOKNLGNNTGDEPCWPCNGSC 171
DB 2530 SDMLCSKGICONTPOPSFTC 2550

```

RESULT 8  
 ADL1\_HUMAN STANDARD: PRT; 769 AA.  
 AC 075078; 014808; 014809; 014810;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)

```

DE ADAM 11 PRECURSOR (A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 11)
DE (METALLOPROTEINASE-LIKE, DISINTEGRIN-LIKE, AND CYSTEINE-RICH PROTEIN)
DE (MDC).
GN ADAM11 OR MDC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=98359734; Pubmed-9693107;
RA Sadane K., Ohya Y., Hasegawa Y., Tanaka I.;
RT "Metalloproteinase-like, disintegrin-like, cysteine-rich proteins MDC2
RT and MDC3: novel human cellular disintegrins highly expressed in the
RT brain.";
RL Biochem. J. 334:93-98(1998).
RN [2]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Cerebellum;
RX MEDLINE=94073190; Pubmed-8252040;
RA Eml M., Katagiri T., Harada Y., Saito H., Inazawa J., Ito I.,
RA Kasumi F., Nakamura Y.;
RT "A novel metalloprotease/disintegrin-like gene at 17q21.3 is
RT somatically rearranged in two primary breast cancers.";
RL Nat. Genet. 5:151-157(1993).
RN [3]
RP SEQUENCE OF 106-769 FROM N.A. (LONG AND SHORT ISOFORMS).
RC TISSUE=Brain, Ovary, Testis, and Breast;
RX MEDLINE=95044425; Pubmed-7956356;
RA Katagiri T., Harada Y., Eml M., Nakamura Y.;
RT "Human metalloprotease/disintegrin-like (MDC) gene: exon-intron
RT organization and alternative splicing.";
RL Cytogenet. Cell Genet. 68:39-44(1995).
CC -1- FUNCTION: PROBABLE LIGAND FOR INTEGRIN IN THE BRAIN. THIS IS A NON
CC CATALYTIC METALLOPROTEASE-LIKE PROTEIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM/MDC-769 (SHOWN HERE)
CC AND A SHORT FORM/MDC-524; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN. SLIGHTLY
CC DEPECTED OR NOT AT ALL IN OTHER TISSUES.
CC -1- DOMAIN: A CONSERVED MOTIF [AVN(E/D)CD] WITHIN THE DISINTEGRIN-LIKE
CC DOMAIN COULD BE INVOLVED IN THE BINDING TO THE INTEGRIN RECEPTOR.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUEN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- DISEASE: REARRANGEMENTS OCCUR IN BREAST AND OVARIAN CANCERS WHICH
CC INVOLVE MULTIPLE EXONS AND DISRUPT THE CODING REGION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AB009675; BAA33352.1; -
DR EMBL: D17390; BAA04213.1; -
DR EMBL: D18172; BAA06670.1; -
DR EMBL: D31872; BAA06671.1; -
DR HSSP: P18619; 1FVL.
DR MEROPS: M12_976; -.
DR TIM: 155120; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR Pfam: PF00200; disintegrin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.

```



SO SEQUENCE 773 AA; 84134 MW; 9A8A125FB3F743DD CRC64;

Query Match 8.4%; Score 82; DB 1; Length 773;  
Best Local Similarity 21.5%; Pred. No. 2.5;  
Matches 45; Conservative 27; Mismatches 61; Indels 76; Gaps 14;

31 LPEICT-----OCPGVONLSKVAFC-----KTRRLMLHA---RC-- 65  
DB 516 IAEPTGDSGSCPPNHLKLD--GYCDEGRGRCGRCKTRDQCALMGHAADRECYE 573  
OY 66 -LMOGKTLIGLONCSLEDPGPNFHOAHT-----TVIIDANPLKDLANTRGFT 117  
DB 574 KLVNGETERG---NCGRK--GSGWVOCSSKODVLCGFLICVNISGAPRLDGLGDISSVT 627  
OY 118 QLOTLILQHVANCPRG--INMAMTITSYID-----NOIC-----Q 150  
DB 628 FYHO--GKELDCRGHVOLDGSDLSVEDGTACGPNMLCLDHRCLPASAFNFTCPGS 684  
OY 151 GOKNLCNNTGDEMPENGSCV--PDGPG 177  
DB 685 GERICSHHG---VCSNKGKICQPDWTG 710

RESULT 10  
FGL2\_MOUSE STANDARD: PRT; 432 AA.  
ID FGL2\_MOUSE  
AC P12804:  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE FIBROBLAST PRECURSOR (FIBRINOGEN-LIKE PROTEIN 2) (PROTHROMBINASE)  
DE (CYTOTOXIC T-LYMPHOCYTE SPECIFIC PROTEIN).  
GN FGL2 OR FIBL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-CYTOTOXIC T-cell;  
RX MEDLINE=87175527; PubMed=3550794;  
RA Koyama T., Hall L.R., Hasegawa S., Saito H.;  
RT Structure of a cytotoxic T-lymphocyte-specific gene shows a strong  
RT homology to fibrinogen beta and gamma chains.  
RL Proc. Natl. Acad. Sci. U.S.A. 84:1609-1613(1987).  
RN [2]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RC STRAIN=BA1B/CJ; TISSUE=Peritoneal macrophage;  
RX MEDLINE=95333285; PubMed=7609073;  
RA Parr R.L., Fung L., Reneker J., Myers-Mason N., Lebowitz J.L.,  
RA Levy G.;  
RT Virus-induced prothrombinase activity.  
RT "Association of mouse fibrinogen-like protein with murine hepatitis  
RT virus-induced prothrombinase activity."  
RL J. Virol. 69:5033-5038(1995).  
CC -1- FUNCTION: CONVERTS PROTHROMBIN TO THROMBIN.  
CC -1- SUBUNIT: HOMOTETRAMER; DISULFIDE-LINKED (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- TISSUE SPECIFICITY: CONSTITUTIVELY EXPRESSED IN CYTOTOXIC  
CC T-CELLS.  
CC -1- INDUCTION: IN MACROPHAGES, DURING INFECTION BY MOUSE HEPATITIS  
CC VIRUS STRAIN 3 (MHV-3).  
CC -1- SIMILARITY: CONTAINS 1 FIBRINOGEN C-TERMINAL DOMAIN.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M16238; AAA37624.1; -

DR EMBL; M15761; AAA37624.1; JOINED.  
DR EMBL; S78773; AAB34823.1; -  
DR PIR; A27447; A27447.  
DR HSSP; P02671; 1FZD.  
DR MCD; MGI:103266; Fg12.  
DR InterPro: IPR002181; Fibrinogen\_C.  
DR Pfam: PF00147; fibrinogen\_C; 1.  
DR SMART: SM00186; FBG; 1.  
DR PROSITE: PS00514; FIBRIN\_AG\_C\_DOMAIN; 1.  
KW T-cell; Cytolysis; Signal.  
FT SIGNAL 1  
FT CHAIN 20 432  
FT DOMAIN 203 428  
FT DISULFID 206 235  
FT DISULFID 364 377  
FT CARBOHYD 24 24  
FT CARBOHYD 172 172  
FT CARBOHYD 228 228  
FT CARBOHYD 256 256  
FT CARBOHYD 329 329  
FT CONFLICT 332 332  
SQ SEQUENCE 432 AA; 48951 MW; 2B297F69CCB4A782 CRC64;  
A -> G (IN REF. 2).

Query Match 8.3%; Score 81; DB 1; Length 432;  
Best Local Similarity 23.2%; Pred. No. 1.6;  
Matches 35; Conservative 22; Mismatches 62; Indels 32; Gaps 6;

OY 29 LALPEICTQCGSVONLSKVAFCYCKTREL--LHARCCLNQKTLIGLONCSLE--- 83  
DB 57 LTLPLTLTLQPLPRLGSMEEVLTKEAVDSLKSC-----QCKKQLAAD 103  
OY 84 --DPGPNFHOAHTV-----IIDLQANPLKDLANTRGFTQLOTLILQHVANCPRG 135  
DB 104 HRDPGNGNGNAGMETEDSRVQLESQVKNLSSELNKAKDQIOGRLTLHL-----VN 158  
OY 136 ANNTITSYDNOICOGKNLICNNTGDEMPCP 166  
DB 159 -MNIENIYVDNIVANLTVVNSLDEKCSKP 188

RESULT 11  
FBN1\_PIG STANDARD: PRT; 2871 AA.  
ID FBN1\_PIG  
AC Q9TV36:  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE FIBRILLIN 1 PRECURSOR.  
DE FBN1.  
GN Sus scrofa (Pig).  
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RX MEDLINE=99156858; PubMed=10036187;  
RA Biery N.J., Eldadah Z.A., Moore C.S., Stetten G., Spencer F.,  
RA Dietz H.C.;  
RT "Revised genomic organization of FBN1 and significance for regulated  
RT gene expression."  
RL Genomics 56:70-77(1999).  
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS  
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE  
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.  
CC -1- PFM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER  
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE  
CC MICROFIBRILS.  
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING  
CC EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.  
CC -----  
CC









	Query March	7.9%;	Score 76.5;	DB 1;	Length 2213;	
	Best Local Similarity	19.2%;	Pred. No. 27;			
	Matches 39;	Conservative 16;	Mismatches 59;	Indels 89;	Gaps 9;	
QY	50	FYCKTFREMLHARCCINCKGTILGIDLCNCLSEDPG-----PNHF-----QAHT 94				
		: :   :	:			

ID	NAME	STANDARD	PRT	886 AA
Db	1280 EVCKNRQOCLFHSWVC	---DGI	---QC	RDSEDEPAFAGCSKDPEFHVKYDDEFGQCON 1333
Qy	95 TVTID	-----QANPLKGLDANFPRG	-----	116
Db	1334 GVCISLIWQCDGDDGSDSEANCENPTEAPNCSTYFQFCRNGHCIPRKKCKDREND	-----	-----	1393
Qy	117	-----TQLQTLILPQHVNCFGINAMVTTSTYIDNOCQKNI	-----	158
Db	1394 GWSDEKDCGSHVLPSTPAPSTCLPNYRCGG	-----ACV	-----	1445
Qy	159 TGDPEMCPE	-----NGSCVPDPG	177	
Db	1446 GSDSEACPSLPNWTATSSPSQPS	1468		
RESULT	13			
EMR1_HUMAN				
ID	EMR1_HUMAN	STANDARD	PRT	886 AA
AC	Q14246			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	20-NOV-2001 (Rel. 40, Last annotation update)			
DE	CELL SURFACE GLYCOPROTEIN EMR1 PRECURSOR (EMR1 HORMONE RECEPTOR).			
GN	EMR1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95324926; PubMed=7601460;			
SA	Roe B.A., Chisoe S.L., Viegas-Pequignot E., Dillon S., Nguyen V.C.,			
RA	Boe B.A., Lipinski M.;			
RT	"EMR1, an unusual member in the family of hormone receptors with			
RL	seven transmembrane segments.";			
CC	Genomics 26:334-344(1995).			
CC	-1- FUNCTION: PROBABLY INVOLVED IN CELLULAR RESPONSE TO A HORMONE OR			
CC	AN INTERACTION WITH A PROTEIN LIGAND.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: WIDE EXPRESSION; INCREASED LEVELS IN			
CC	PERIPHERAL BLOOD MONONUCLEAR CELLS.			
CC	-1- PTM: N- AND O-GLYCOSYLATED; (POSSIBLE).			
CC	-1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>			
CC	or send a request to <a href="mailto:isb@isb.ch">isb@isb.ch</a> ).			
CC	-----			
DR	EMBL: X81117; YAA57232.1; -.			
DR	HSSP: P000008; -.			
DR	MM: 60045; -.			
DR	InterPro: IPR000152; Asx_hydroxyl.			
DR	InterPro: IPR000561; EGF-like.			
DR	InterPro: IPR001881; EGF_Ca.			
DR	InterPro: IPR000832; GPCR_secretin.			
DR	InterPro: IPR000203; PKC_cys_rich.			
DR	Pfam: PF000002; 7tm_2; 1.			
DR	Pfam: PF00008; EGF_5.			
DR	Pfam: PF01825; GPS; 1.			
DR	PRINTS: PRO1128; EMR1HORMONER.			
DR	SMART: SM00179; EGF_CA; 5.			
DR	SMART: SM00303; GPS; 1.			
DR	PROSITE: PS00650; G_PROTEIN_RECP_F2_2; 1.			
DR	PROSITE: PS00010; ASX_HYDROXYL; 6.			
DR	PROSITE: PS01166; EGF_2; 2.			
DR	PROSITE: PS01187; EGF_CA; 5.			
DR	G-protein coupled receptor; Transmembrane; Receptor; Glycoprotein;			

Db 201 REFESSGHLSGGLKASCDIEDCTEMCPINSTCT-NTPG 240

RESULT 14

ID	KAL_HUMAN	STANDARD:	PRT:	638 AA.
AC	P03952:			
DT	23-OCT-1986 (Rel. 02, Created)			
DT	23-OCT-1986 (Rel. 02, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)			
DE	(KININOGENIN) (FLETCHER FACTOR).			
GN	KLK3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RA	SEQUENCE FROM N.A.			
RA	MEDLINE=8624359; PubMed=3521732;			
RA	Chung D.W., Fujikawa K., McMullen B.A., Davie E.W.;			
RT	"Human plasma prekallikrein, a zymogen to a serine protease that			
RT	contains four tandem repeats.";			
RL	Biochemistry 25:2410-2417(1986).			
RN	[2]			
RN	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RA	MEDLINE=91152016; PubMed=1998666;			
RA	McMullen B.A., Fujikawa K., Davie E.W.;			
RT	"Location of the disulfide bonds in human plasma prekallikrein: the			
RT	presence of four novel apple domains in the amino-terminal portion of			
RT	the molecule.";			
RL	Biochemistry 30:2050-2056(1991).			
CC	-1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT			
CC	ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING			
CC	TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM			
CC	HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN			
CC	SYSTEM BY CONVERTING PRORENIN INTO RENIN.			
CC	-1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES			
CC	THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,			
CC	AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE			
CC	CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.			
CC	-1- DISEASE: DEFECTS IN KLK3 ARE THE CAUSE OF FLETCHER FACTOR			
CC	DEFICIENCY, A BLOOD COAGULATION DEFECT.			
CC	-1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on lists			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL; M13143; AAA60153.1; -.			
DR	PIR; A00921; KOHUP.			
DR	PIR; A37939; A37939.			

```

DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390
FT CHAIN 391 638
FT DOMAIN 20 105
FT DOMAIN 110 195
FT DOMAIN 200 285
FT DOMAIN 291 376
FT DOMAIN 389 621
FT CARBOHYD 127 127
FT CARBOHYD 308 308
FT CARBOHYD 396 396
FT CARBOHYD 453 453
FT CARBOHYD 494 494
FT ACT_SITE 434 434
FT ACT_SITE 483 483
FT ACT_SITE 578 578
FT ACT_SITE 21 104
FT DISULFID 47 77
FT DISULFID 51 57
FT DISULFID 111 134
FT DISULFID 137 166
FT DISULFID 141 147
FT DISULFID 201 284
FT DISULFID 227 256
FT DISULFID 231 237
FT DISULFID 292 375
FT DISULFID 318 347
FT DISULFID 322 328
FT DISULFID 340 345
FT DISULFID 383 503
FT DISULFID 419 435
FT DISULFID 517 584
FT DISULFID 548 563
FT DISULFID 574 602
SO SEQUENCE 638 AA; 71369 MW; E62F9C1053838FB4 CRC64;

Query Match 7.6%; Score 74; DB 1; Length 638;
Best Local Similarity 19.4%; Pred. NO. 12;
Matches 45; Conservative 25; Mismatches 40; Indels 122; Gaps 13;

OY 14 PMAALLALGVERALLALPEICTQCPG-----VONT 45
DB 403 PMOVSIAQVKLTAQRHL-----CGSLIGHQWVLAHCFDGLPLQDVWRITSGILNL 454
OY 46 SKVAFYCKT-----RELMLARCCLNQKTIILGLDQNCSLDDPGFNHQAHTTVIIDL 100
DB 455 SDIT---KDFPSQIKETIIL-----QNKVSEGN---HDIALIKL 489
OY 101 QAN-----PLKGLDANFR-----GFT---QLOTLILPOHVNCPGGINMW 137
DB 490 QAPLWTEFQKICLPKSGDSTIYTNCWVTGWFSGEKEIIONIL--QKVNIP----- 541
OY 138 NTITSYDNOICQ-----GOKNLC--NNTGDPENCPENG 169
DB 542 -----LVITNECOKRYODYKITQRMWCAGYKGGKDACKDSGGLPYCKHNG 588

RESULT 15
PUT3_YEAST STANDARD; PRT; 979 AA.
AC P25302;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROLINE UTILIZATION TRANS-ACTIVATOR.
FT METAL 34 60
FT METAL 37 37
OS Saccharomyces cerevisiae (Baker's yeast).

```

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91203881; PubMed=2017167;
RA Marczak J.E., Brandtiss M.C.;
RT "Analysis of constitutive and noninducible mutations of the PUT3
RT transcriptional activator.";
RN Mol. Cell. Biol. 11:2609-2619(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=94205264; PubMed=8154185;
RA Wiemann S., Voss H., Schwager C., Rupp T., Stegemann J.,
RA Zimmermann J., Grothues D., Sensen C., Erfle H., Hewitt N.,
RA Baurevi A., Ansorge W.;
RT "Sequencing and analysis of 51.6 kilobases on the left arm of
RT chromosome XI from Saccharomyces cerevisiae reveals 23 open reading
RT frames including the FAS1 gene.";
RN yeast 9:1343-1348(1993).
RN [3]
RP SEQUENCE OF 1-825 FROM N.A.
RA Rieger M.;
RT Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP STRUCTURE BY NMR OF 30-100.
RA MEDLINE=97448678; PubMed=9303003;
RA Walters K.J., Dayle K.T., Reece R.J., Ptashne M., Wagner G.;
RT "Structure and mobility of the PUT3 dimer.";
RN Nat. Struct. Biol. 4:744-750(1997).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 31-100.
RA MEDLINE=97448679; PubMed=9303004;
RA Swaminathan K., Flynn P., Reece R.J., Marmorstein R.;
RT "Crystal structure of a PUT3-DNA complex reveals a novel mechanism
RT for DNA recognition by a protein containing a Zn2Cys6 binuclear
RT cluster.";
RN Nat. Struct. Biol. 4:751-759(1997).
CC -1- FUNCTION: POSITIVE ACTIVATOR OF THE PROLINE UTILIZATION PATHWAY.
CC BINDS TO THE PROMOTERS OF PUT1 AND PUT2 GENES. RECOGNIZES AND
CC BINDS TO THE DNA SEQUENCE 5'-CGG-N(10)-CCG-3'.
CC -1- SUBUNIT: BINDS DNA AS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is provided through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Users may and for commercial
CC entities must obtain a license agreement (See http://www.isb-sib.ch/announce/
CC or send e-mail to license@sib-sib.ch).
CC
DR EMBL; X5; AA39055.1; -
DR EMBL; X7; AA52267.1; -
DR EMBL; Z284; AA81850.1; -
DR PIR; A39753; A39792.
DR PDB; 1AJY; 1-SEP-97.
DR PDB; 1ZME; 1-SEP-98.
DR TRANSFAC; TC 163; -
DR SGD; S0001498; PUT3.
DR InterPro; IPR001138; ZN2_CYS6_fungal.
DR Pfam; PF00174; Zn_Cls1; 1.
DR PRINTS; PR00054; FUNGALZNCYS.
DR SMART; SM00666; GAL4; 1.
DR PROSITE; PS00463; ZN2_CYS6_FUNGAL_1; 1.
DR PROSITE; PS50048; ZN2_CYS6_FUNGAL_2; 1.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Zinc; Metal-binding; Proline metabolism; 3D-structure;
FT DNA_BIND 34 60
FT METAL 34 34
FT METAL 37 37
FT METAL 37 37
ZINC 1.
ZINC 1 AND 2.
ZINC 1.

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:25:36 ; Search time 21.01 seconds  
(without alignments)  
308,885 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_1\_177

Perfect score: 177  
Sequence: 1 MAPHGPGSLTTLVPMMAALL.....NTGDPEMCPEMSCVDPDGP 177

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1802

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5	2.8	15	1	ASP1_LACSN
2	5	2.8	25	1	ALR_PSEFL
3	5	2.8	30	1	ACON_CANAL
4	5	2.8	28	1	VAAL_EQUAR
5	4	2.3	10	1	TKSI_AEDAE
6	4	2.3	12	1	FREI_LITIN
7	4	2.3	13	1	CHRP_PARID
8	4	2.3	15	1	KLOM_LOMTE
9	4	2.3	15	1	PKRH_PHYPA
10	4	2.3	16	1	CERB_RAT
11	4	2.3	16	1	PX_RAT
12	4	2.3	17	1	BOL5_MEGPE
13	4	2.3	18	1	DRPH_UCAPU
14	4	2.3	18	1	TKN2_SCYCA
15	4	2.3	19	1	ITHA_PERAM
16	4	2.3	19	1	MDH_SHEPU
17	4	2.3	20	1	CP35_PAPSP
18	4	2.3	20	1	CPBX_CAVPO
19	4	2.3	20	1	MIF_PIG
20	4	2.3	21	1	EPa8_HUMAN
21	4	2.3	21	1	MDH_KLEPN
22	4	2.3	22	1	ATP6_COTJA
23	4	2.3	22	1	LPT6_SERMA
24	4	2.3	23	1	CR41_LITCE
25	4	2.3	23	1	CR42_LITCE
26	4	2.3	23	1	CR43_LITCE
27	4	2.3	23	1	GP50_BESPI
28	4	2.3	23	1	PRP1_RAT
29	4	2.3	24	1	LPER_STRFR
30	4	2.3	24	1	PSAC_CUCSA
31	4	2.3	24	1	RAN_XENLA
32	4	2.3	25	1	ACP_ERYLO
33	4	2.3	25	1	ATP0_SPTOL

34	4	2.3	25	1	ATPD_MICLU	P80285 micrococcus
35	4	2.3	25	1	COXO_ONCMY	P80334 oncorhynch
36	4	2.3	25	1	METE_CANAL	P82610 candida alb
37	4	2.3	25	1	PA22_BONSC	P80963 bothriechis
38	4	2.3	25	1	RS19_ACHLA	P29224 acholeplasm
39	4	2.3	25	1	RT06_BOVIN	P82931 bos taurus
40	4	2.3	25	1	UBU1_BOVIN	P23356 bos taurus
41	4	2.3	26	1	COX2_SOLITU	P80498 solanum tub
42	4	2.3	27	1	PA59_CAPIH	P80934 capra hircu
43	4	2.3	28	1	ARV4_PSEPU	P35902 pseudomonas
44	4	2.3	29	1	AL21_HORSE	P81216 equus cabal
45	4	2.3	29	1	PSAX_SYNVU	P23320 synechococc

## ALIGNMENTS

RESULT 1  
ID ASP1\_LACSN STANDARD; PRT; 15 AA.  
AC P82648:  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DE ACID SHOCK PROTEIN 1 (FRAGMENT).  
OS Lactobacillus sanfranciscensis (Lactobacillus sanfrancisco).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Lactobacillus;  
OX NCBI\_TaxID=1625;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=CB1;  
RX PubMed=11429463;  
RA De Angelis M., Bini L., Pallini V., Coccocelli P.S., Gobbetti M.;  
RT "The acid-stress response in Lactobacillus sanfranciscensis CBL";  
RL Microbiology 147:1863-1873(2001).  
CC -i- INDUCTION: OVEREXPRESSED IN ACID ENVIRONMENT.  
FT NON\_TER 15 15  
SQ SEQUENCE 15 AA; 1509 MW; 575853B4DFB030A8 ...4;

Query Match 2.8%; Score 5; DB 1;  
Best Local 5; Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 GTTCT  
DB 9 GTTCT

RESULT 2  
ID ALR\_PSEFL STANDARD; PRT; 25 AA.  
AC P33162:  
DT 01-OCT-1991 (Rel. 1, Created)  
DT 01-OCT-1991 (Rel. 1, Last sequence update)  
DE 01-FEB-1992 (Rel. 28, Last annotation update)  
DE ALANINE AMYLASE (EC 5.1.1.1) (FRAGMENT).  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas;  
OX NCBI\_TaxID=2294;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=TW5-2;  
RX MEDLINE=93169026; PubMed=7763424;  
RA Yokoigawa K., Kawai H., Iim Y.H., Esaki N., Soda K.;  
RT "Thermolabile alanine racemase from a psychrophilic Pseudomonas  
fluorescens: purification and properties";  
RL Biosci. Biotechnol. Biochem. 57:93-97(1993).  
CC -i- FUNCTION: CELL WALL FORMATION. IT IS THERMOLABILE AND SHOWS HIGH  
ACTIVITY AT LOW TEMPERATURES.  
CC -i- CATALYTIC ACTIVITY: L-ALANINE = D-ALANINE.

CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: ALONG WITH D-ALANINE-D-ALANINE LIGASE, IT MAKES UP THE  
 CC D-ALANINE BRANCH OF THE PEPTIDOGLYCAN BIOSYNTHETIC ROUTE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SIMILARITY: BELONGS TO THE ALANINE RACEMASE FAMILY.  
 DR PIR: PC1221; PC1221.  
 DR InterPro: IPR000821; Ala racemase.  
 DR PROSITE: PS00385; ALANINE\_RACEMASE; PARTIAL.  
 KW Isomerase; Pyridoxal phosphate; Cell wall; Peptidoglycan synthesis.  
 FT NON\_TER 25  
 FT 25  
 SQ SEQUENCE 25 AA; 2774 MW; 54D5109F62B1A43 CRC64;

Query Match 2.8%; Score 5; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 8 IDLQA 102  
 8 IDLQA 12

RESULT 3  
 ACON\_CANAL STANDARD: PRT: 28 AA.  
 AC P82611;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ACONITATE HYDRAE, MITOCHONDRIAL (EC 4.2.1.3) (CITRATE HYDRO-LYASE)  
 DE (ACONITASE) (FRAGMENTS).  
 GN ACO1.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE. AND MASS SPECTROMETRY.  
 RC SRRAIN-SC5314;  
 RX MEDLINE=20403418; PubMed=10949142;  
 RA Pardo M., Ward M., Platch A., Sanchez M., Nombela C., Blackstock W.,  
 RA Gil C.;  
 RT "Cross-species identification of novel Candida albicans immunogenic  
 RT proteins by combination of two-dimensional polyacrylamide gel  
 RT electrophoresis and mass spectrometry.";  
 RL Electrophoresis 21:2651-2659(2000).  
 CC -1- CATALYTIC ACTIVITY: CITRATE - CIS-ACONITATE + H(2)O.  
 CC -1- COFACTOR: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE (3FE-  
 CC 4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE CATALYTIC  
 CC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND ISOCITRATE (BY  
 CC SIMILARITY).  
 CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.  
 CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.  
 DR InterPro: IPR001030; Aconitase.  
 DR PROSITE: PS00450; ACONITASE\_1; PARTIAL.  
 DR PROSITE: PS00450; ACONITASE\_2; PARTIAL.  
 KW Lyase; Tricarboxylic acid cycle; Iron-sulfur; Mitochondrion; 4Fe-4S.  
 FT NON\_TER 1  
 FT 1  
 FT METAL 7  
 FT 7  
 FT UNSURE 3  
 FT 3  
 FT UNSURE 4  
 FT 4  
 FT NON\_CONS 17  
 FT 17  
 FT NON\_TER 28  
 FT 28  
 SQ SEQUENCE 28 AA; 3036 MW; C8D118B97D071291 CRC64;

Query Match 2.8%; Score 5; DB 1; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 PGSVQ 43

Db 22 PGSVQ 26

RESULT 4  
 VAAL\_EQUAR STANDARD: PRT: 30 AA.  
 ID VAAL\_EQUAR  
 AC 004236;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT A, ISOFORM 1 (EC 3.6.1.34)  
 DE (FRAGMENT).  
 OS Equisetum arvense (Field horsetail) (Common horsetail).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Equisetophyta; Sphenopsida; Equisetales; Equisetaceae; Equisetum.  
 OX NCBI\_TaxID=3258;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93138084; PubMed=8422915;  
 RA Starke T., Gogarten J.P.;  
 RT "A conserved intron in the V-ATPase A subunit genes of plants and  
 RT algae.";  
 RL FEBS Lett. 315:252-258(1993).  
 CC -1- FUNCTION: CATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF  
 CC VACUOLAR ATPASE. V-ATPASE VACUOLAR ATPASE IS RESPONSIBLE FOR  
 CC ACIDIFYING A VARIETY OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC  
 CC CELLS.  
 CC -1- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A  
 CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,  
 CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE  
 CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).  
 CC -1- MISCELLANEOUS: TWO SEPARATE GENES ENCODE THE CATALYTIC 70 KDA  
 CC V-ATPASE SUBUNIT IN PSILLOTUM AND EUISETUM.  
 CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: X56983; CAA40301.1; -;  
 DR InterPro: IPR00194; ATPase\_alpha\_beta.  
 DR PROSITE: PS00152; ATPASE\_ALPHA\_BETA; PARTIAL.  
 KW ATP synthetase; Hydrogen ion transport; Hydrolase; ATP-binding;  
 KW Multigene family.  
 FT NON\_TER 1  
 FT 1  
 FT NON\_TER 30  
 FT 30  
 SQ SEQUENCE 30 AA; 3337 MW; 9627ED62068D761F CRC64;

Query Match 2.8%; Score 5; DB 1; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LITLV 13  
 25 LITLV 29

RESULT 5  
 TKSL\_AEDAE STANDARD: PRT: 10 AA.  
 ID TKSL\_AEDAE  
 AC P42634;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE SIALOKININ I.  
 OS Aedes aegypti (Yellowfever mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Aedes.  
 OX NCBI\_TaxID=7159;  
 RN [1]  
 RP SEQUENCE  
 RC STRAIN-ROCKEFELLER; TISSUE-Salivary gland;  
 RA MEDLINE=94105119; PubMed=8278354;  
 RT Champagne D.E., Ribeiro J.M.C.;  
 RT "Stalokinin I and II: vasodilatory tachykinins from the yellow fever  
 mosquito *Aedes aegypti*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:138-142(1994).  
 CC -1- FUNCTION: VASODILATORY PEPTIDE. MAY ACTIVATE MACROPHAGES AT THE  
 CC SITE OF FEEDING.  
 CC -1- SIMILARITY: SOME SIMILARITY TO TACHYKININS.  
 DR PIR: A49581; A49581.  
 DR InterPro: IPR002040; Tachykinin.  
 DR PROSITE: PS00267; TACHYKININ; 1.  
 KW Tachykinin; Neuropeptide; Amidation.  
 FT MOD\_RES 10 10 AMIDATION.  
 SQ SEQUENCE 10 AA; 1145 MW; 3DCFD6B59C33A8 CRC64;

Query Match 2.3%; Score 4; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 158 NTGD 161  
 DB 1 NTGD 4

RESULT 6  
 PREL\_LITIN STANDARD; PRT; 12 AA.  
 AC P82021;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PRENATIN 1.  
 OS Litoria infrafrenata (Giant tree frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 OC Litoria.  
 OX NCBI\_TaxID=61195;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC TISSUE-Skin secretion;  
 RX MEDLINE=97368637; PubMed=9225251;  
 RA Raftery M.J., Maugh R.J., Bowie J.H., Wallace J.C., Tyler M.J.;  
 RT "The structures of the frenatin peptides from the skin secretion of  
 the giant tree frog *Litoria infrafrenata*.";  
 RL J. Pept. Sci. 2:117-124(1996).  
 CC -1- FUNCTION: WIDE SPECTRUM ANTIMICROBIAL PEPTIDE.  
 CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN PAROTOID AND/OR ROSTRAL  
 CC GLANDS.  
 CC -1- MASS SPECTROMETRY: MW=1140; METHOD=FAE.  
 CC Antibiotic; Amphibian skin; Amidation.  
 KW MOD\_RES 12 12 AMIDATION.  
 FT SEQUENCE 12 AA; 1141 MW; C622550BC365B72D CRC64;

Query Match 2.3%; Score 4; DB 1; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 ILGL 75  
 DB 9 ILGL 12

RESULT 7  
 CHEP\_PARID STANDARD; PRT; 13 AA.  
 ID CHEP\_PARID

AC P42718;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE CHEMOTACTIC PEPTIDE.  
 OS Parapolybia indica.  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
 OC Vespoidea; Vespidae; Polistinae; Parapolybia.  
 OX NCBI\_TaxID=31921;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=venom;  
 RA Toki T., Yasunara T., Nakajima T.;  
 RT "Isolation and sequential analysis of peptides on the venom sac of  
*Parapolybia indica*.";  
 RL Eisai Dobutsu 39:105-111(1988).  
 KW Chemotaxis; Amidation.  
 FT MOD\_RES 13 13 AMIDATION.  
 SQ SEQUENCE 13 AA; 1298 MW; 5C950CE8E39D5873 CRC64;

Query Match 2.3%; Score 4; DB 1; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 ILGL 75  
 DB 1 ILGL 4

RESULT 8  
 KLOM\_LUMTE STANDARD; PRT; 15 AA.  
 AC P11918;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE LOMBRIKINE KINASE (EC 2.7.3.5) (LK) (FRAGMENT).  
 OS Lumbricus terrestris (Common earthworm).  
 OC Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
 OC Lumbricina; Lumbricidae; Lumbricus.  
 OX NCBI\_TaxID=709;  
 RN [1]  
 RP SEQUENCE  
 RC MEDLINE=720665; PubMed=5128744;  
 RA det Terronstan E., Desvages G., Pradel L.A., N. R., van Thoi N.;  
 RT "Comparative structural studies of the active site of ATP: guanidine  
 phosphotransferase. The essential cysteine tripeptide of  
 lumbric kinase from *Lumbricus terrestris* murex."  
 RL Eur. J. Biochem. 22:585-592(1971).  
 CC -1- CATALYTIC ACTIVITY: ATP + LOMBRIKINE = ADP + PHOSPHOLOMBRIKINE.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SIMILARITY: BELONGS TO THE ATP:GUANIDINO PHOSPHOTRANSFERASE FAMILY.  
 DR PIR: A08; A4416.  
 DR InterPro: PRO00749; ATP\_gua\_Ptrans.  
 DR PROSITE: PS00112; GUANIDINO\_KINASE; 1.  
 KW Kinase; Transferase.  
 FT NON\_TER 1 1  
 FT ACT\_SITE 6 6  
 FT NON\_TER 15 15  
 SQ SEQUENCE 15 AA; 1565 MW; 2A45FE6F6140E90C4 CRC64;

Query Match 2.3%; Score 4; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 CPGS 42  
 DB 6 CPGS 9

```

RESULT 9
PKKH_PHYPA STANDARD; PRT: 15 AA.
ID PKKH_PHYPA
AC P80659;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHOSPHOGLYCERATE KINASE, CHLOROPLAST (EC 2.7.2.3) (FRAGMENT).
OS Physcomitrella patens (Moss).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Bryopsida; Funariaceae; Funariales; Funariaceae; Physcomitrella.
OX NCBI_TaxID=3218;
RN [1]
RP SEQUENCE.
RC TISSUE=Protonema; PubMed=9129336;
RX MEDLINE=97275459;
RA Kasten B., Buck F., Nuske J., Reski R.;
RT "Cytoxinin affects nuclear- and plastome-encoded energy-converting
RT plastid enzymes.";
RL Planta 201:261-272(1997).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROL, PHOSPHATE.
CC -1- PATHWAY: CALVIN CYCLE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
DR InterPro: IPR001576; PKR.
DR PROSITE: PS00111; GLYCERATE_KINASE; PARTIAL.
KM Transferase; Kinase; Multigene family; Calvin cycle; Chloroplast.
FT NON_TER 15
SQ SEQUENCE 15 AA: 1531 MW: 56A5BC1F677EE66 CRC64;

Query Match 2.3%; Score 4; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 SKVA 49
DB 7 SKVA 10

RESULT 10
CERB_RAT STANDARD; PRT: 16 AA.
ID CERB_RAT
AC P23436;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CEREBELLIN.
OS Rattus norvegicus (Rat), and Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116, 9823;
RN [1]
RP SEQUENCE.
RC SPECIES=rat;
RX Blacher R., Danho W., Hempstead J.L., Morgan J.I.;
RA "Isolation and sequencing of two cerebellum-specific peptides.";
RT Proc. Natl. Acad. Sci. U.S.A. 81:6866-6870(1984).
RL [2]
RN [1]
RP SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=89341798; PubMed=2760624;
RA Ylangou Y., Burnet P., Nikou G., Chrysanthou B.J., Bloom S.R.;
RT "Purification and characterization of cerebellins from human and
RT porcine cerebellum.";
RL J. Neurochem. 53:886-889(1989).
CC -1- FUNCTION: CEREBELLIN EXERTS NEUROMODULATORY FUNCTIONS. DIRECTLY
CC STIMULATES NOREPINEPHRINE RELEASE VIA THE ADENYLATE CYCLASE/PKA-
CC DEPENDENT SIGNALING PATHWAY. AND INDIRECTLY ENHANCES
CC ADRENOCHORTICAL SECRETION IN VIVO, THROUGH A PARACRINE MECHANISM
CC INVOLVING MEDULLARY CATECHOLAMINE RELEASE.

```

```

CC -1- TISSUE SPECIFICITY: LOCALIZED IN THE PURKINJE CELLS.
DR PIR: A03135; CORF.
DR PIR: PLO124; PLO124.
KM Synapsosome.
SQ SEQUENCE 16 AA: 1633 MW: 3EPAL63343D518 CRC64;

Query Match 2.3%; Score 4; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 KVAF 50
DB 5 KVAF 8

RESULT 11
PX_RAT STANDARD; PRT: 16 AA.
ID PX_RAT
AC P14630;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE PROTEIN PX (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RX MEDLINE=90122905; PubMed=2297521;
RA Blatter M.-C., James R.W., Borghini I., Martin B.M.,
RA Hochstrasser A.-C., Pomella D.;
RT "A novel high-density lipoprotein particle and associated protein in
RT rat plasma.";
RL Biochim. Biophys. Acta 1042:19-27(1990).
CC -1- FUNCTION: UNKNOWN.
CC -1- TISSUE SPECIFICITY: PLASMA PROTEIN ASSOCIATED WITH HDL.
KM Plasma; HDL.
FT NON_TER 16
SQ SEQUENCE 16 AA: 1740 MW: 79836ECDAE9FBA29 CRC64;

Query Match 2.3%; Score 4; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.9e+03;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LTTL 12
DB 12 LTTL 15

RESULT 12
BOLS_MEGPE STANDARD; PRT: 17 AA.
ID BOLS_MEGPE
AC P07496;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)
DE BOMBOLITIN V.
OS Megabombus pennsylvanicus (American common bumblebee).
OC Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Apocrita; Aculeata;
OC Apoidea; Apidae; Bombus.
OX NCBI_TaxID=28643;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=85105003; PubMed=2578459;
RA Argiolas A., Pisano J.J.;
RT "Bombolitins, a new class of mast cell degranulating peptides from
RT the venom of the bumblebee Megabombus pennsylvanicus.";
RL J. Biol. Chem. 260:1437-1444(1985).

```



CC -1- FUNCTION: MAST CELL DEGRANULATING PEPTIDE.  
 DR PIR: E22595; E22595.  
 KW Mast cell degranulation; Venom.  
 SQ SEQUENCE 17 AA; 1731 MW; B076C091571606A5 CRC64;

Query Match 2.3%; Score 4; DB 1; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 2e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 ILGL 75  
 ||||  
 Db 6 ILGL 9

RESULT 13  
 DRPH\_UCAPU STANDARD; PRT; 18 AA.  
 AC P08871;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PIGMENT-DISPERSING HORMONE (PDH) (LIGHT ADAPTING DISTAL RETINAL  
 DE PIGMENT HORMONE) (DRPH).  
 OS Uca pugilator (Atlantic sand fiddler crab) (Celuca pugilator).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Ocypodidae; Ocypodinae; Uca complex;  
 OC Celuca  
 OX NCBI\_TaxID=6772;  
 RN [1]  
 RP SEQUENCE.  
 RA Rao K.R., Riehm J.P., Zahnow C.A., Kleinholz L.H., Tarr G.E.,  
 RA Johnson L., Norton S., Landau M., Semmes O.J., Sattelberg R.M.,  
 RA Jorenbey W.H., Hintz M.F.;  
 RT "Characterization of a pigment-dispersing hormone in eyestalks of the  
 RT fiddler crab Uca pugilator.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:5319-5322(1985).  
 RN [2]  
 RP SEQUENCE.  
 RX MEDLINE=93230895; PubMed=8472537;  
 RA Loehr J., Klein J., Webster S.G., Dirksen H.;  
 RT "Quantification, immunofluorescence purification and sequence analysis of  
 RT a pigment-dispersing hormone of the shore crab, Carcinus maenas  
 RT (L.).";  
 RL Comp. Biochem. Physiol. 104B:699-706(1993).  
 CC -1- FUNCTION: CAUSES THE MIGRATION OF THE DISTAL RETINAL PIGMENT  
 CC INTO THE PROXIMAL END OF THE PIGMENT CHROMATOPHORE CELLS AND  
 CC THUS DECREASES THE AMOUNT OF LIGHT ENTERING THE RETINULAS.  
 CC -1- SIMILARITY: TO THE PDH OF OTHER ARTHROPODS.  
 CC PIR: A25144; DRPPD.  
 DR Hormone; Amidation.  
 KW DOMAIN  
 FT MOD\_RES 6 9 IMPORTANT FOR DRPH ACTIVITY.  
 FT 18 18 AMIDATION.  
 SQ SEQUENCE 18 AA; 1928 MW; 25DCE8D016F544E CRC64;

Query Match 2.3%; Score 4; DB 1; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 ILGL 75  
 ||||  
 Db 8 ILGL 11

RESULT 14  
 TKN2\_SCYCA STANDARD; PRT; 18 AA.  
 AC P08609;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE SCYLIORHININ II (RECTIN).  
 OS Scylliorhinus canicula (Spotted dogfish) (Spotted catshark).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 OC Elasmobranchii; Galeomorphii; Galeidae; Carcharhiniformes;  
 OC Scylliorhinidae; Scylliorhinus.  
 OX NCBI\_TaxID=7830;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=86192829; PubMed=2422058;  
 RA Conlon J.M., Deacon C.F., O'Toole L., Thim L.;  
 RT "Scylliorhinin I and II: two novel tachykinins from dogfish gut.";  
 RL FEBS Lett. 200:111-116(1986).  
 RN [2]  
 RP SEQUENCE.

RC TISSUE=Small Intestine;  
 RX MEDLINE=95335921; PubMed=7541963;  
 RA Anderson W.G., Conlon J.M., Hazen N.;  
 RT "Characterization of the endogenous intestinal peptide that  
 RT stimulates the rectal gland of Scylliorhinus canicula.";  
 RL Am. J. Physiol. 268:R1359-R1364(1995).

CC -1- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,  
 CC EVOKE BEHAVIORAL RESPONSES, ARE POTENT VASODILATORS AND  
 CC SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH  
 CC MUSCLES.  
 CC -1- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.

DR PIR: B24867; B24867.  
 DR InterPro: IPR002040; Tachykinin.  
 DR Pfam: PF02202; Tachykinin; 1.  
 DR PROSITE: PS00267; TACHYKININ; 1.  
 KW Tachykinin; Neuropeptide; Amidation.  
 FT DISULFID 7 13  
 FT MOD\_RES 18 18 AMIDATION.  
 SQ SEQUENCE 18 AA; 1854 MW; FCCA3FEB01E2F3907 CRC64;

Query Match 2.3%; Score 4; DB 1; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 173 PDGP 17A  
 ||||  
 Db 8 PDGP 11

RESULT 15  
 ITHA\_PERAM STANDARD; PRT; 19 AA.  
 AC P19986;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE TREHALASE INHIBITOR (PFRAGMENT).  
 OS Periplaneta americana (American cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 OC Blattodea; Blattellidae; Periplaneta.  
 OX NCBI\_TaxID=6978;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Hemolymph;  
 RX MEDLINE=89380218; PubMed=2777784;  
 RA Hayakawa Y., Jaggiardar A.P., Yaguchi M., Downer R.G.H.;  
 RT "Purification and characterization of trehalase inhibitor from  
 RT hemolymph of the American cockroach, Periplaneta americana.";  
 RL J. Biol. Chem. 264:16165-16169(1989).  
 CC -1- FUNCTION: UNDER RESTING CONDITIONS, INHIBITS TREHALASE IN A  
 CC DOSE-DEPENDENT MANNER.  
 DR PIR: A34233; A34233.  
 DR Glycoprotein.  
 KW NON\_TER 19  
 FT SEQUENCE 19 AA; 2016 MW; A60C3A3397AF9A22 CRC64;

Thu Mar 28 09:21:41 2002

us-09-726-348-2\_copy\_1\_177.rsp

Page 6

Query Match 2.3%; Score 4; DB 1; Length 19;  
Best Local Similarity 100.0%; Pred. No. 2.2e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 172 VPDG 175  
1111  
Db 12 VPDG 15

Search completed: March 28, 2002, 09:25:36  
Job time: 534 sec

Thu May 28 09:21:39 2002

us-09-726-348-2\_copy\_126\_177\_1.isp

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model  
March 28, 2002, 09:14:51 ; Search time 62.93 Seconds  
(without alignments)  
Run on: 30.297 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_126\_177  
Sequence: 1 QHVNCPGGINAMNTITSYID.....NTGDEMCPCNGSCVPDGRG 52  
Perfect score: 309

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	67	21.7	2019 1	CIN5_RAT
2	66.5	21.5	1820 1	CIN5_RAT
3	63.5	20.6	886 1	EMRL_HUMAN
4	63	20.4	2476 1	ZAN_PIG
5	63	20.4	2871 1	FBN1_HUMAN
6	62.5	20.2	1964 1	MYC4_MOUSE
7	61.5	19.9	746 1	GNL1_STRE
8	61.5	19.9	2005 1	CIN2_MOUSE
9	61.5	19.9	2907 1	FBN2_MOUSE
10	61.5	19.9	2907 1	FBN2_MOUSE
11	60.5	19.6	2005 1	FBN2_MOUSE
12	60.5	19.6	2871 1	FBN1_BOVIN
13	60.5	19.6	2911 1	FBN2_HUMAN
14	60	19.4	48 1	TX26_PMONI
15	60	19.4	279 1	COAT_SBMV
16	60	19.4	1951 1	CIN3_HUMAN
17	60	19.4	1951 1	CIN3_HUMAN
18	60	19.4	2871 1	FBN1_MOUSE
19	59.5	19.3	495 1	Y243_AOUAE
20	59.5	19.3	568 1	STP_CHIMP
21	59.5	19.3	1353 1	VCU2_CVHOC
22	59.5	19.3	2871 1	FBN1_PIG
23	59.5	19.3	49 1	TX25_PMONI
24	59.5	19.3	1155 1	ROM1_YEAST
25	59.5	19.3	1522 1	CIN1_LODBL
26	59.5	19.3	1840 1	YLR3_CAEEL
27	58.5	18.8	116 1	CIN4_RAT
28	58.5	18.8	2016 1	MCS_HUMAN
29	57.5	18.6	1836 1	CIN5_HUMAN
30	57.5	18.6	1836 1	CIN4_HUMAN
31	57	18.4	161 1	RNKO_GORGO
32	57	18.4	161 1	RNKO_HUMAN
33	57	18.4	161 1	RNKO_PANTR

RESULT	1	STANDARD:	PRT:	2019 AA.	ALIGNMENTS
CIN5_RAT					
AC	P15389				
DT	01-APR-1990 (Rel. 14, last sequence update)				
DT	01-APR-1990 (Rel. 14, last sequence update)				
DT	15-DEC-1998 (Rel. 37, last annotation update)				
DE	SODIUM CHANNEL PROTEIN, CARDIAC MUSCLE ALPHA-SUBUNIT.				
GN	SCNSA:				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
RP	(1)				
RC	SEQUENCE FROM N.A.				
RC	TISSUE=heart;				
RC	MEDLINE=90046760; PubMed=2554302;				
RA	Rogart R.B., Cribbs L.H., Muglia L.K., Kephart D.D., Kaiser M.W.;				
RT	"Molecular cloning of a putative tetrodotoxin-resistant rat heart Na+ channel isoform."				
RT	Proc. Natl. Acad. Sci. U.S.A. 86:8170-8174(1989).				
RL	Proc. Natl. Acad. Sci. U.S.A. 86:8170-8174(1989).				
CC	-1- FUNCTIONALITY: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT SODIUM ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE MEMBRANE. THE PROTEIN FORMS A SODIUM-SELECTIVE CHANNEL THROUGH WHICH NA+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL GRADIENT. IT IS A TETRODOTOXIN-RESISTANT NA+ CHANNEL ISOFORM.				
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				
CC	-1- DOMAIN: THE SEQUENCE CONTAINS 4 INTERNAL REPEATS, EACH WITH 5 HYDROPHOBIC SEGMENTS (S1,S2,S3,S5,S6) AND ONE POSITIVELY CHARGED SEGMENT (S4). SEGMENTS S4 ARE PROBABLY THE VOLTAGE-SENSORS AND ARE CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.				
CC	-1- MISCELLANEOUS: NA+ CHANNELS IN MAMMALIAN CARDIAC MEMBRANE HAVE FUNCTIONAL PROPERTIES QUITE DISTINCT FROM NA+ CHANNELS IN NERVE AND SKELETAL MUSCLE.				
CC	-1- SIMILARITY: TO OTHER SODIUM CHANNEL PROTEINS.				
CC	-1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb.slb.ch">license@isb.slb.ch</a> ).				
CC	EMBL: M27902; AAA2114.1;				
CC	PIR: A33996; A33996.				
DR	InterPro: IPR002111; Cat_channel_TyPL.				
DR	InterPro: IPR000636; Cation_chan_not_11g.				
DR	InterPro: IPR001682; Channel_pore_Ca_Na.				
DR	InterPro: IPR000048; IO.				
DR	InterPro: IPR001696; Na_channel.				
DR	InterPro: IPR00520; Ion_trans_4.				
DR	Pfam: PF00612; IO_1.				
DR	PRINTS: PR00170; NACCHANNEL.				

Q08731 mus musculus  
Q05209 o solitio-  
Q41435 solanum tub  
P47783 macaca fasc  
P00763 rattus norv  
P13837 paramecium  
P54704 dictyostel  
Q09115 brachiocto  
P35498 homo sapien  
P04774 rattus norv  
Q61549 mus musculu  
Q9m081 arabidopsi

DR SMART: SM00015; IO: 1.  
 KW Glycine channel: Transmembrane; Ion transport: Voltage-gated channel;  
 FT TRANSMEM 126 151 S1 OF REPEAT I (POTENTIAL).  
 FT TRANSMEM 160 179 S2 OF REPEAT I (POTENTIAL).  
 FT TRANSMEM 193 211 S3 OF REPEAT I (POTENTIAL).  
 FT TRANSMEM 218 237 S4 OF REPEAT I (POTENTIAL).  
 FT TRANSMEM 254 277 S5 OF REPEAT I (POTENTIAL).  
 FT TRANSMEM 391 416 S6 OF REPEAT I (POTENTIAL).  
 FT TRANSMEM 713 737 S7 OF REPEAT I (POTENTIAL).  
 FT TRANSMEM 749 772 S1 OF REPEAT II (POTENTIAL).  
 FT TRANSMEM 781 800 S2 OF REPEAT II (POTENTIAL).  
 FT TRANSMEM 807 826 S3 OF REPEAT II (POTENTIAL).  
 FT TRANSMEM 843 863 S4 OF REPEAT II (POTENTIAL).  
 FT TRANSMEM 917 942 S5 OF REPEAT II (POTENTIAL).  
 FT TRANSMEM 1203 1226 S6 OF REPEAT II (POTENTIAL).  
 FT TRANSMEM 1240 1265 S2 OF REPEAT III (POTENTIAL).  
 FT TRANSMEM 1272 1293 S3 OF REPEAT III (POTENTIAL).  
 FT TRANSMEM 1298 1319 S4 OF REPEAT III (POTENTIAL).  
 FT TRANSMEM 1339 1361 S5 OF REPEAT III (POTENTIAL).  
 FT TRANSMEM 1446 1472 S6 OF REPEAT III (POTENTIAL).  
 FT TRANSMEM 1526 1549 S1 OF REPEAT IV (POTENTIAL).  
 FT TRANSMEM 1561 1584 S2 OF REPEAT IV (POTENTIAL).  
 FT TRANSMEM 1591 1614 S3 OF REPEAT IV (POTENTIAL).  
 FT TRANSMEM 1625 1646 S4 OF REPEAT IV (POTENTIAL).  
 FT TRANSMEM 1662 1684 S5 OF REPEAT IV (POTENTIAL).  
 FT TRANSMEM 1750 1774 S6 OF REPEAT IV (POTENTIAL).  
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 549 549 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 741 741 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 804 804 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 842 842 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 865 865 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1367 1367 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1376 1376 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1382 1382 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1390 1390 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 2019 AA; 227366 MW; CFC3B03CEAE708AD CRC64.

Query Match 21.7%; Score 67; DB 1; Length 2019;  
 Best Local Similarity 29.6%; Pred. No. 11;  
 Matches 16; Conservative 8; Mismatches 20; Indels 10; Gaps 2;

OY 7 GINA-----WNTTSTIDNQCQGRN-----LCNNTGPEKCPENGSCVDPG 50  
 DB 293 GSVADGLVWNSLDVYLDNPANYLLKNTTYVLLCGNSSDAGTCPEGYCLNAG 346

RESULT 2  
 CINA\_ELEEL STANDARD; PRT; 1820 AA.  
 AC P02719;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DE 15-JUL-1998 (Rel. 36, Last annotation update)  
 OS SODIUM CHANNEL PROTEIN (NA+ CHANNEL).  
 OS Electrophorus electricus (Electric eel).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Gymnalliformes; Electrophoridae; Electrophorus.  
 NCBI\_TaxID=8005;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85061498; PubMed=6209577;  
 RA Node M., Shimizu S., Tanabe T., Takai T., Kayano T., Ikeda T.,

RA Takahashi H., Nakayama H., Kanaoka Y., Miamino N., Kangawa K.,  
 RA Matsuo H., Kakeyama M.A., Hirose T., Inayama S., Hayashida H.,  
 RA Miyata T., Numa S.;  
 RT "Primary structure of Electrophorus electricus sodium channel deduced  
 RT from cDNA sequence.";  
 RL Nature 312:121-127(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87311395; PubMed=2442385;  
 RA Node M., Numa S.;  
 RT "Structure and function of sodium channel.";  
 RL J. Recept. Res. 7:467-497(1987).  
 CC -1- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT SODIUM ION  
 CC PERMEABILITY OF EXCITABLE MEMBRANES, ASSUMING OPENED OR CLOSED  
 CC MEMBRANE, THE PROTEIN FORMS A SODIUM-SELECTIVE CHANNEL THROUGH  
 CC WHICH NA+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL  
 CC GRADIENT.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- DOMAIN: THE SEQUENCE CONTAINS 4 INTERNAL REPEATS, EACH WITH 5  
 CC HYDROPHOBIC SEGMENTS (S1, S2, S3, S5, S6) AND ONE POSITIVELY CHARGED  
 CC SEGMENT (S4). SEGMENTS S4 ARE PROBABLY THE VOLTAGE-SENSORS AND ARE  
 CC CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT  
 CC EVERY THIRD POSITION.  
 CC -1- MISCELLANEOUS: AVAILABLE DATA SUGGEST THAT ACTIVATION AND  
 CC INACTIVATION GATES ARE LOCATED NEAR THE CYTOPLASMIC SURFACE OF THE  
 CC MEMBRANE. IT IS HYPOTHEZIZED THAT RESIDUES 802-806, 847-857,  
 CC 894-910, AND 942-955 MIGHT, IN CONJUNCTION WITH THE POSITIVELY  
 CC CHARGED RESIDUES OF S4, ACT AS A VOLTAGE SENSOR INVOLVED WITH THE  
 CC ACTIVATION GATE.  
 CC -1- SIMILARITY: TO OTHER SODIUM CHANNEL PROTEINS.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch).)  
 DR EMBL: X01119; CAA25587.1; -;  
 DR EMBL: M22522; AAA79960.1; -;  
 DR PIR: A03178; CHEE.  
 DR InterPro: IPR002113; Cat\_channel\_Trl.  
 DR InterPro: IPR000636; Catlon\_chan\_non\_lig.  
 DR InterPro: IPR001682; Channel\_pore\_Ca\_Na.  
 DR Pfam: PF00520; Ion\_trans\_4.  
 DR PRINTS: PR00170; NACHANNEL.  
 DR KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
 KW Glycoprotein; Repeat.  
 FT REPEAT 111 419 I.  
 FT REPEAT 555 807 II.  
 FT REPEAT 989 1281 III.  
 FT REPEAT 1311 1587 IV.  
 FT TRANSMEM 118 138 S1 OF REPEAT I.  
 FT TRANSMEM 150 171 S2 OF REPEAT I.  
 FT TRANSMEM 177 197 S3 OF REPEAT I.  
 FT TRANSMEM 204 224 S4 OF REPEAT I.  
 FT TRANSMEM 244 264 S5 OF REPEAT I.  
 FT TRANSMEM 285 342 NON-HOMOLOGOUS REGION OF REPEAT I.  
 FT TRANSMEM 379 402 S6 OF REPEAT I.  
 FT TRANSMEM 558 578 S1 OF REPEAT II.  
 FT TRANSMEM 600 620 S2 OF REPEAT II.  
 FT TRANSMEM 626 643 S3 OF REPEAT II.  
 FT TRANSMEM 651 671 S4 OF REPEAT II.  
 FT TRANSMEM 691 711 S5 OF REPEAT II.  
 FT TRANSMEM 767 790 S6 OF REPEAT II.  
 FT TRANSMEM 1005 1025 S1 OF REPEAT III.  
 FT TRANSMEM 1038 1058 S2 OF REPEAT III.  
 FT TRANSMEM 1066 1086 S3 OF REPEAT III.  
 FT TRANSMEM 1092 1112 S4 OF REPEAT III.  
 FT TRANSMEM 1132 1152 S5 OF REPEAT III.

FT DOMAIN 1172 1194 NON-HOMOLOGOUS REGION OF REPEAT III.  
 FT TRANSMEM 1244 1264 S6 OF REPEAT III.  
 FT TRANSMEM 1321 1341 S1 OF REPEAT IV.  
 FT TRANSMEM 1353 1376 S2 OF REPEAT IV.  
 FT TRANSMEM 1381 1398 S3 OF REPEAT IV.  
 FT TRANSMEM 1417 1437 S4 OF REPEAT IV.  
 FT TRANSMEM 1454 1474 S5 OF REPEAT IV.  
 FT DOMAIN 1490 1505 NON-HOMOLOGOUS REGION OF REPEAT IV.  
 FT TRANSMEM 1544 1567 S6 OF REPEAT IV.  
 FT CARBOHYD 278 288 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 690 690 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 797 797 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1160 1160 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1820 AA; 208331 MW; 1B271F626E057864 CRC64;

Query Match 21.5%; Score 66.5; DB 1; Length 1820;  
 Best Local Similarity 34.0%; Pred. No. 11;  
 Matches 17; Conservative 8; Mismatches 16; Indels 9; Gaps 3;

Oy 10 NAMNT---ITSYIDNQ---ICQGQKN-LCNGTGDPEKPCENSCVPDG 50  
 Db 285 SAVNTFTFTAYIENENQYFLDGLDALLCGNSDAGCPGYCYCMAG 334

RESULT 3  
 ID EMR1\_HUMAN STANDARD; PRT; 886 AA.

AC 014246;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CELL SURFACE GLYCOPROTEIN EMR1 PRECURSOR (EMR1 HORMONE RECEPTOR).  
 GN EMR1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95324926; PubMed=7601460;  
 RA Baud V., Chissac S.L., Viegas-Pequignot E., Diliong S., N'Guyen V.C.,  
 RA Roe B.A., Lipinski M.;  
 RT \*EMR1, an unusual member in the family of hormone receptors with  
 seven transmembrane segments.\*;  
 RL Genomics 26:334-344(1995).  
 CC -!- FUNCTION: PROBABLY INVOLVED IN CELLULAR RESPONSE TO A HORMONE OR  
 AN INTERACTION WITH A PROTEIN LIGAND.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- TISSUE SPECIFICITY: WIDE EXPRESSION; INCREASED LEVELS IN  
 PERIPHERAL BLOOD MONONUCLEAR CELLS.  
 CC -!- PTM: N- AND O-GLYCOSYLATED; (POSSIBLE).  
 CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X81479; CAA57232.1; -  
 DR HSSP: P00736; IAPQ.  
 DR MIM: 600493; -  
 DR Interpro: IPR000152; Asx\_hydroxyl.  
 DR Interpro: IPR000561; EGF-like.  
 DR Interpro: IPR001881; EGF\_Ca.

DR Interpro: IPR000832; GPCR\_secretin.  
 DR Interpro: IPR000203; PKD\_cys\_rich.  
 DR Pfam: PR00002; 7tm\_2; 1.  
 DR Pfam: PF00008; EGF; 5.  
 DR Pfam: PF01825; GPs; 1.  
 DR PRINTS: PR01128; EMR1HORMONER.  
 DR SMART: SM00179; EGF\_CA; 5.  
 DR SMART: SM00001; EGF\_1like; 1.  
 DR SMART: SM00303; GPs; 1.  
 DR PROSITE: PS00650; G-PROTEIN\_RECPT\_F2\_2; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 6.  
 DR PROSITE: PS01186; EGF\_2; 2.  
 DR PROSITE: PS01187; EGF\_CA; 5.  
 KW G-protein coupled receptor; Transmembrane; Receptor; Glycoprotein;  
 FT SIGNAL 1 17  
 FT CHAIN 18 886  
 FT DOMAIN 18 599  
 FT TRANSMEM 600 627  
 FT DOMAIN 628 634  
 FT TRANSMEM 635 656  
 FT DOMAIN 657 666  
 FT TRANSMEM 667 690  
 FT DOMAIN 691 709  
 FT TRANSMEM 710 731  
 FT DOMAIN 732 747  
 FT TRANSMEM 748 776  
 FT DOMAIN 777 794  
 FT TRANSMEM 795 814  
 FT DOMAIN 815 829  
 FT TRANSMEM 830 852  
 FT DOMAIN 853 886  
 FT DOMAIN 31 79  
 FT DOMAIN 80 131  
 FT DOMAIN 132 171  
 FT DOMAIN 172 213  
 FT DOMAIN 221 267  
 FT DOMAIN 268 316  
 FT DOMAIN 317 599  
 FT DISULFID 35 47  
 FT DISULFID 41 56  
 FT DISULFID 58 78  
 FT DISULFID 84 97  
 FT DISULFID 91 106  
 FT DISULFID 108 130  
 FT DISULFID 136 148  
 FT DISULFID 142 157  
 FT DISULFID 159 170  
 FT DISULFID 176 188  
 FT DISULFID 182 197  
 FT DISULFID 199 212  
 FT DISULFID 225 235  
 FT DISULFID 229 244  
 FT DISULFID 246 266  
 FT DISULFID 272 285  
 FT DISULFID 279 294  
 FT DISULFID 296 315  
 FT CARBOHYD 94 94  
 FT CARBOHYD 99 99  
 FT CARBOHYD 127 127  
 FT CARBOHYD 167 167  
 FT CARBOHYD 189 189  
 FT CARBOHYD 194 194  
 FT CARBOHYD 232 232  
 FT CARBOHYD 258 258  
 FT CARBOHYD 312 312  
 FT CARBOHYD 366 366  
 FT CARBOHYD 375 375  
 FT CARBOHYD 448 448  
 FT CARBOHYD 661 661  
 SQ SEQUENCE 886 AA; 97680 MW; 7456CA36F8624D99 CRC64;



- OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=93372860; Pubmed=8364578;  
 RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,  
 RA Panglisan T., Bonadio J.;  
 RT "Genomic organization of the sequence coding for fibrillin, the  
 RT defective gene product in Marfan syndrome.";  
 RL Hum. Mol. Genet. 2:961-968(1993).  
 [2]  
 RP SEQUENCE OF 1-932 FROM N.A.  
 RC TISSUE=Placenta, and Fibroblast;  
 RX MEDLINE=94010947; Pubmed=7691719;  
 RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;  
 RT "Fibrillin binds calcium and is coded by cDNAs that reveal a  
 RT multidomain structure and alternatively spliced exons at the 5'  
 RT end.";  
 RL Genomics 17:476-484(1993).  
 [3]  
 RP SEQUENCE OF 899-2871 FROM N.A.  
 RX MEDLINE=91304568; Pubmed=1852207;  
 RA Maslen C.L., Corson G.M., Maddox B.K., Glanville R.W., Sakai L.Y.;  
 RT "Partial sequence of a candidate gene for the Marfan syndrome.";  
 RL Nature 352:334-337(1991).  
 [4]  
 RP SEQUENCE OF 813-1313 FROM N.A.  
 RX MEDLINE=91304567; Pubmed=1852206;  
 RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,  
 RA Tsipouras P., Ramirez F., Hollister D.W.;  
 RT "Linkage of Marfan syndrome and a phenotypically related disorder to  
 RT two different fibrillin genes.";  
 RL Nature 352:330-334(1991).  
 [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=91317849; Pubmed=1860873;  
 RA Sakai L.Y., Keene D.R., Glanville R.W., Bachinger H.P.;  
 RT "Purification and partial characterization of fibrillin, a cysteine-  
 RT rich structural component of connective tissue microfibrils.";  
 RL J. Biol. Chem. 266:14763-14770(1991).  
 [6]  
 RP STRUCTURE BY NMR OF 2054-2125.  
 RX MEDLINE=98031893; Pubmed=9362480;  
 RA Yuan X., Downing A.K., Knott V., Handford P.A.;  
 RT "Solution structure of the transforming growth factor beta-binding  
 RT protein-like module, a domain associated with matrix fibrils.";  
 RL EMBO J. 16:6659-6666(1997).  
 [7]  
 RP STRUCTURE BY NMR OF 2124-2205.  
 RX MEDLINE=9614829; Pubmed=8568869;  
 RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;  
 RT "Calcium binding properties of an epidermal growth factor-like domain  
 RT pair from human fibrillin-1.";  
 RL J. Mol. Biol. 255:22-27(1996).  
 [8]  
 RP STRUCTURE BY NMR OF 2124-2205.  
 RX MEDLINE=9622301; Pubmed=8653794;  
 RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,  
 RA Handford P.A.;  
 RT "Solution structure of a pair of calcium-binding epidermal growth  
 RT factor-like domains: implications for the Marfan syndrome and other  
 RT genetic disorders.";  
 RL Cell 85:597-605(1996).  
 [9]  
 RP REVIEW ON MFS VARIANTS.  
 RX MEDLINE=96174615; Pubmed=8594563;  
 RA Collod G., Beroud C., Soussi T., Junten C., Boileau C.;  
 RT "Software and database for the analysis of mutations in the human  
 RT FBN1 gene.";  
 RL Nucleic Acids Res. 24:137-141(1996).  
 [10]  
 RP REVIEW ON MFS VARIANTS.  
 RX MEDLINE=97169383; Pubmed=9016526;  
 RA Collod-Beroud G., Beroud C., Ades L., Black C., Boxer M., Brock D.J.,  
 RA Godfrey M., Hayward C., Karttunen L., Milewicz D., Peltonen L.,  
 RA Richards R.J., Wang W., Junten C., Boileau C.,  
 RT "Marfan Database (second edition): software and database for the  
 RT analysis of mutations in the human FBN1 gene.";  
 RL Nucleic Acids Res. 25:147-150(1997).  
 [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=98062175; Pubmed=9401003;  
 RA Hayward C., Brock D.J.H.;  
 RT "Fibrillin-1 mutations in Marfan syndrome and other type-1  
 RT fibrillinopathies.";  
 RL Hum. Mutat. 10:415-423(1997).  
 [12]  
 RP VARIANT MFS PRO-1137.  
 RX MEDLINE=91304569; Pubmed=1852208;  
 RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,  
 RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,  
 RA Christin S.M., Stetten G., Meyers D.A., Francomano C.A.;  
 RT "Marfan syndrome caused by a recurrent de novo missense mutation in  
 RT the fibrillin gene.";  
 RL Nature 352:337-339(1991).  
 [13]  
 RP VARIANTS MFS SER-1249, ARG-1663, SER-2221 AND SER-2307.  
 RX MEDLINE=93250834; Pubmed=1301946;  
 RA Dietz H.C., Saraiya J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;  
 RT "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome  
 RT patients at cysteine residues in EGF-like domains.";  
 RL Hum. Mutat. 1:366-374(1992).  
 [14]  
 RP VARIANT MFS SER-2307.  
 RX MEDLINE=92235290; Pubmed=1569206;  
 RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J., Jr.,  
 RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;  
 RT "Marfan phenotype variability in a family segregating a missense  
 RT mutation in the epidermal growth factor-like motif of the fibrillin  
 RT gene.";  
 RL J. Clin. Invest. 89:1674-1680(1992).  
 [15]  
 RP VARIANTS MFS ILE-548 AND ALA-723.  
 RX MEDLINE=94010946; Pubmed=8406497;  
 RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,  
 RA Pyeritz R.E., Francomano C.A.;  
 RT "Four novel FBN1 mutations: significance for mutant transcript level  
 RT and EGF-like domain calcium binding in the pathogenesis of Marfan  
 RT syndrome.";  
 RL Genomics 17:468-475(1993).  
 [16]  
 RP VARIANTS MFS SER-2144.  
 RX MEDLINE=93278402; Pubmed=8504310;  
 RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;  
 RT "A novel fibrillin mutation in the Marfan syndrome which could  
 RT disrupt calcium binding of the epidermal growth factor-like module.";  
 RL Hum. Mol. Genet. 2:475-477(1993).  
 [17]  
 RP VARIANTS MFS R-862, Y-1117, P-1137 AND F-1589, AND VARIANT A-1148.  
 RX MEDLINE=94108431; Pubmed=8281141;  
 RA Tynan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,  
 RA Berg M.A., Miller D.C., Francke U.;  
 RT "Mutation screening of complete fibrillin-1 coding sequence: report  
 RT of five new mutations, including two in 8-cysteine domains.";  
 RL Hum. Mol. Genet. 2:1813-1821(1993).  
 [18]  
 RP VARIANTS MFS GLY-217 AND ARG-2627.  
 RX MEDLINE=95067970; Pubmed=797366;  
 RA Karttunen L., Reghunath M., Loenqvist L., Peltonen L.;  
 RT "A compound-heterozygous Marfan patient: two defective fibrillin  
 RT alleles result in a lethal phenotype.";  
 RL Am. J. Hum. Genet. 55:1083-1091(1994).  
 [19]  
 RP VARIANT EL LYS-2447.





```

FT DOMAIN 311 350 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 352 388 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 389 427 EGF-LIKE 10.
FT DOMAIN 429 470 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 472 508 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 510 546 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 548 584 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 586 622 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 622 656 EGF-LIKE 16.
FT DOMAIN 658 686 EGF-LIKE 17.
FT DOMAIN 688 724 EGF-LIKE 18.
FT DOMAIN 726 762 EGF-LIKE 19.
FT DOMAIN 764 800 EGF-LIKE 20.
FT DOMAIN 803 839 EGF-LIKE 21.
FT DOMAIN 841 877 EGF-LIKE 22.
FT DOMAIN 878 924 EGF-LIKE 23.
FT DOMAIN 926 962 EGF-LIKE 24.
FT DOMAIN 964 1000 EGF-LIKE 25.
FT DOMAIN 1002 1040 EGF-LIKE 26.
FT DOMAIN 1042 1081 EGF-LIKE 27.
FT DOMAIN 1083 1122 EGF-LIKE 28.
FT DOMAIN 1126 1167 EGF-LIKE 29.
FT REPEAT 1168 1208 LIN/NOTCH 1.
FT REPEAT 1209 1242 LIN/NOTCH 2.
FT REPEAT 1243 1282 LIN/NOTCH 3.
FT REPEAT 1282 1321 ANK 1.
FT REPEAT 1321 1360 ANK 2.
FT REPEAT 1360 1400 ANK 3.
FT REPEAT 1400 1440 ANK 4.
FT REPEAT 1440 1480 ANK 5.
FT REPEAT 1480 1520 ANK 6.
FT REPEAT 1520 1560 ANK 7.
FT REPEAT 1560 1600 ANK 8.
FT REPEAT 1600 1640 ANK 9.
FT REPEAT 1640 1680 ANK 10.
FT REPEAT 1680 1720 ANK 11.
FT REPEAT 1720 1760 ANK 12.
FT REPEAT 1760 1800 ANK 13.
FT REPEAT 1800 1840 ANK 14.
FT REPEAT 1840 1880 ANK 15.
FT REPEAT 1880 1920 ANK 16.
FT REPEAT 1920 1960 ANK 17.
FT REPEAT 1960 2000 ANK 18.
FT REPEAT 2000 2040 ANK 19.
FT REPEAT 2040 2080 ANK 20.
FT REPEAT 2080 2120 ANK 21.
FT REPEAT 2120 2160 ANK 22.
FT REPEAT 2160 2200 ANK 23.
FT REPEAT 2200 2240 ANK 24.
FT REPEAT 2240 2280 ANK 25.
FT REPEAT 2280 2320 ANK 26.
FT REPEAT 2320 2360 ANK 27.
FT REPEAT 2360 2400 ANK 28.
FT REPEAT 2400 2440 ANK 29.
FT REPEAT 2440 2480 ANK 30.
FT REPEAT 2480 2520 ANK 31.
FT REPEAT 2520 2560 ANK 32.
FT REPEAT 2560 2600 ANK 33.
FT REPEAT 2600 2640 ANK 34.
FT REPEAT 2640 2680 ANK 35.
FT REPEAT 2680 2720 ANK 36.
FT REPEAT 2720 2760 ANK 37.
FT REPEAT 2760 2800 ANK 38.
FT REPEAT 2800 2840 ANK 39.
FT REPEAT 2840 2880 ANK 40.
FT REPEAT 2880 2920 ANK 41.
FT REPEAT 2920 2960 ANK 42.
FT REPEAT 2960 3000 ANK 43.
FT REPEAT 3000 3040 ANK 44.
FT REPEAT 3040 3080 ANK 45.
FT REPEAT 3080 3120 ANK 46.
FT REPEAT 3120 3160 ANK 47.
FT REPEAT 3160 3200 ANK 48.
FT REPEAT 3200 3240 ANK 49.
FT REPEAT 3240 3280 ANK 50.
FT REPEAT 3280 3320 ANK 51.
FT REPEAT 3320 3360 ANK 52.
FT REPEAT 3360 3400 ANK 53.
FT REPEAT 3400 3440 ANK 54.
FT REPEAT 3440 3480 ANK 55.
FT REPEAT 3480 3520 ANK 56.
FT REPEAT 3520 3560 ANK 57.
FT REPEAT 3560 3600 ANK 58.
FT REPEAT 3600 3640 ANK 59.
FT REPEAT 3640 3680 ANK 60.
FT REPEAT 3680 3720 ANK 61.
FT REPEAT 3720 3760 ANK 62.
FT REPEAT 3760 3800 ANK 63.
FT REPEAT 3800 3840 ANK 64.
FT REPEAT 3840 3880 ANK 65.
FT REPEAT 3880 3920 ANK 66.
FT REPEAT 3920 3960 ANK 67.
FT REPEAT 3960 4000 ANK 68.
FT REPEAT 4000 4040 ANK 69.
FT REPEAT 4040 4080 ANK 70.
FT REPEAT 4080 4120 ANK 71.
FT REPEAT 4120 4160 ANK 72.
FT REPEAT 4160 4200 ANK 73.
FT REPEAT 4200 4240 ANK 74.
FT REPEAT 4240 4280 ANK 75.
FT REPEAT 4280 4320 ANK 76.
FT REPEAT 4320 4360 ANK 77.
FT REPEAT 4360 4400 ANK 78.
FT REPEAT 4400 4440 ANK 79.
FT REPEAT 4440 4480 ANK 80.
FT REPEAT 4480 4520 ANK 81.
FT REPEAT 4520 4560 ANK 82.
FT REPEAT 4560 4600 ANK 83.
FT REPEAT 4600 4640 ANK 84.
FT REPEAT 4640 4680 ANK 85.
FT REPEAT 4680 4720 ANK 86.
FT REPEAT 4720 4760 ANK 87.
FT REPEAT 4760 4800 ANK 88.
FT REPEAT 4800 4840 ANK 89.
FT REPEAT 4840 4880 ANK 90.
FT REPEAT 4880 4920 ANK 91.
FT REPEAT 4920 4960 ANK 92.
FT REPEAT 4960 5000 ANK 93.
FT REPEAT 5000 5040 ANK 94.
FT REPEAT 5040 5080 ANK 95.
FT REPEAT 5080 5120 ANK 96.
FT REPEAT 5120 5160 ANK 97.
FT REPEAT 5160 5200 ANK 98.
FT REPEAT 5200 5240 ANK 99.
FT REPEAT 5240 5280 ANK 100.
FT REPEAT 5280 5320 ANK 101.
FT REPEAT 5320 5360 ANK 102.
FT REPEAT 5360 5400 ANK 103.
FT REPEAT 5400 5440 ANK 104.
FT REPEAT 5440 5480 ANK 105.
FT REPEAT 5480 5520 ANK 106.
FT REPEAT 5520 5560 ANK 107.
FT REPEAT 5560 5600 ANK 108.
FT REPEAT 5600 5640 ANK 109.
FT REPEAT 5640 5680 ANK 110.
FT REPEAT 5680 5720 ANK 111.
FT REPEAT 5720 5760 ANK 112.
FT REPEAT 5760 5800 ANK 113.
FT REPEAT 5800 5840 ANK 114.
FT REPEAT 5840 5880 ANK 115.
FT REPEAT 5880 5920 ANK 116.
FT REPEAT 5920 5960 ANK 117.
FT REPEAT 5960 6000 ANK 118.
FT REPEAT 6000 6040 ANK 119.
FT REPEAT 6040 6080 ANK 120.
FT REPEAT 6080 6120 ANK 121.
FT REPEAT 6120 6160 ANK 122.
FT REPEAT 6160 6200 ANK 123.
FT REPEAT 6200 6240 ANK 124.
FT REPEAT 6240 6280 ANK 125.
FT REPEAT 6280 6320 ANK 126.
FT REPEAT 6320 6360 ANK 127.
FT REPEAT 6360 6400 ANK 128.
FT REPEAT 6400 6440 ANK 129.
FT REPEAT 6440 6480 ANK 130.
FT REPEAT 6480 6520 ANK 131.
FT REPEAT 6520 6560 ANK 132.
FT REPEAT 6560 6600 ANK 133.
FT REPEAT 6600 6640 ANK 134.
FT REPEAT 6640 6680 ANK 135.
FT REPEAT 6680 6720 ANK 136.
FT REPEAT 6720 6760 ANK 137.
FT REPEAT 6760 6800 ANK 138.
FT REPEAT 6800 6840 ANK 139.
FT REPEAT 6840 6880 ANK 140.
FT REPEAT 6880 6920 ANK 141.
FT REPEAT 6920 6960 ANK 142.
FT REPEAT 6960 7000 ANK 143.
FT REPEAT 7000 7040 ANK 144.
FT REPEAT 7040 7080 ANK 145.
FT REPEAT 7080 7120 ANK 146.
FT REPEAT 7120 7160 ANK 147.
FT REPEAT 7160 7200 ANK 148.
FT REPEAT 7200 7240 ANK 149.
FT REPEAT 7240 7280 ANK 150.
FT REPEAT 7280 7320 ANK 151.
FT REPEAT 7320 7360 ANK 152.
FT REPEAT 7360 7400 ANK 153.
FT REPEAT 7400 7440 ANK 154.
FT REPEAT 7440 7480 ANK 155.
FT REPEAT 7480 7520 ANK 156.
FT REPEAT 7520 7560 ANK 157.
FT REPEAT 7560 7600 ANK 158.
FT REPEAT 7600 7640 ANK 159.
FT REPEAT 7640 7680 ANK 160.
FT REPEAT 7680 7720 ANK 161.
FT REPEAT 7720 7760 ANK 162.
FT REPEAT 7760 7800 ANK 163.
FT REPEAT 7800 7840 ANK 164.
FT REPEAT 7840 7880 ANK 165.
FT REPEAT 7880 7920 ANK 166.
FT REPEAT 7920 7960 ANK 167.
FT REPEAT 7960 8000 ANK 168.
FT REPEAT 8000 8040 ANK 169.
FT REPEAT 8040 8080 ANK 170.
FT REPEAT 8080 8120 ANK 171.
FT REPEAT 8120 8160 ANK 172.
FT REPEAT 8160 8200 ANK 173.
FT REPEAT 8200 8240 ANK 174.
FT REPEAT 8240 8280 ANK 175.
FT REPEAT 8280 8320 ANK 176.
FT REPEAT 8320 8360 ANK 177.
FT REPEAT 8360 8400 ANK 178.
FT REPEAT 8400 8440 ANK 179.
FT REPEAT 8440 8480 ANK 180.
FT REPEAT 8480 8520 ANK 181.
FT REPEAT 8520 8560 ANK 182.
FT REPEAT 8560 8600 ANK 183.
FT REPEAT 8600 8640 ANK 184.
FT REPEAT 8640 8680 ANK 185.
FT REPEAT 8680 8720 ANK 186.
FT REPEAT 8720 8760 ANK 187.
FT REPEAT 8760 8800 ANK 188.
FT REPEAT 8800 8840 ANK 189.
FT REPEAT 8840 8880 ANK 190.
FT REPEAT 8880 8920 ANK 191.
FT REPEAT 8920 8960 ANK 192.
FT REPEAT 8960 9000 ANK 193.
FT REPEAT 9000 9040 ANK 194.
FT REPEAT 9040 9080 ANK 195.
FT REPEAT 9080 9120 ANK 196.
FT REPEAT 9120 9160 ANK 197.
FT REPEAT 9160 9200 ANK 198.
FT REPEAT 9200 9240 ANK 199.
FT REPEAT 9240 9280 ANK 200.
FT REPEAT 9280 9320 ANK 201.
FT REPEAT 9320 9360 ANK 202.
FT REPEAT 9360 9400 ANK 203.
FT REPEAT 9400 9440 ANK 204.
FT REPEAT 9440 9480 ANK 205.
FT REPEAT 9480 9520 ANK 206.
FT REPEAT 9520 9560 ANK 207.
FT REPEAT 9560 9600 ANK 208.
FT REPEAT 9600 9640 ANK 209.
FT REPEAT 9640 9680 ANK 210.
FT REPEAT 9680 9720 ANK 211.
FT REPEAT 9720 9760 ANK 212.
FT REPEAT 9760 9800 ANK 213.
FT REPEAT 9800 9840 ANK 214.
FT REPEAT 9840 9880 ANK 215.
FT REPEAT 9880 9920 ANK 216.
FT REPEAT 9920 9960 ANK 217.
FT REPEAT 9960 10000 ANK 218.

```

```

FT DISULFID 595 610 BY SIMILARITY.
FT DISULFID 612 621 BY SIMILARITY.
FT DISULFID 626 637 BY SIMILARITY.
FT DISULFID 631 646 BY SIMILARITY.
FT DISULFID 648 655 BY SIMILARITY.
FT DISULFID 662 669 BY SIMILARITY.
FT DISULFID 664 674 BY SIMILARITY.
FT DISULFID 666 676 BY SIMILARITY.
FT DISULFID 676 685 BY SIMILARITY.
FT DISULFID 682 703 BY SIMILARITY.
FT DISULFID 697 712 BY SIMILARITY.
FT DISULFID 714 723 BY SIMILARITY.
FT DISULFID 730 741 BY SIMILARITY.
FT DISULFID 735 750 BY SIMILARITY.
FT DISULFID 752 761 BY SIMILARITY.
FT DISULFID 768 779 BY SIMILARITY.
FT DISULFID 773 788 BY SIMILARITY.
FT DISULFID 790 799 BY SIMILARITY.
FT DISULFID 807 818 BY SIMILARITY.
FT DISULFID 812 827 BY SIMILARITY.
FT DISULFID 829 838 BY SIMILARITY.
FT DISULFID 845 856 BY SIMILARITY.
FT DISULFID 850 865 BY SIMILARITY.
FT DISULFID 867 876 BY SIMILARITY.
FT DISULFID 882 903 BY SIMILARITY.
FT DISULFID 897 912 BY SIMILARITY.
FT DISULFID 914 923 BY SIMILARITY.
FT DISULFID 930 941 BY SIMILARITY.
FT DISULFID 935 950 BY SIMILARITY.
FT DISULFID 952 961 BY SIMILARITY.
FT DISULFID 968 979 BY SIMILARITY.
FT DISULFID 973 988 BY SIMILARITY.
FT DISULFID 990 999 BY SIMILARITY.
FT DISULFID 1006 1019 BY SIMILARITY.
FT DISULFID 1011 1028 BY SIMILARITY.
FT DISULFID 1030 1039 BY SIMILARITY.
FT DISULFID 1046 1057 BY SIMILARITY.
FT DISULFID 1051 1069 BY SIMILARITY.
FT DISULFID 1071 1080 BY SIMILARITY.
FT DISULFID 1087 1098 BY SIMILARITY.

Query Match 20.2% Score 62.5; DB 1; Length 1964;
Best Local Similarity 30.9%; Pred. No. 32;
Matches 17; Conservative 7; Mismatches 18; Indels 13; Gaps 3;

QY 5 CPGGIMANTITSY-----IDNIOGQGNKNCNNTGDEPKPENGSC--VPDG 50
Db 202 CPGGTSCHNLTGSYQCLCPVGQEGPQCKLRKAC---PGSCLNGGTCLVPEG 252

RESULT 7
GUNI_STRRE STANDARD: PRT: 746 AA.
ID GUNI_STRRE
AC Q05156;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE CELLULOSE 1 PRECURSOR (EC 3.2.1.4) (ENDOGLUCANASE) (ENDO-1,4-BETA-
DE GLUCANASE) (AVICELASE).
GN CELL.
OS Streptomyces reticuli.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1926;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-41 AND 379-391.
RC STRAIN=TU45; PubMed=1282194;
RX MEDLINE=93116600; PubMed=1282194;
RA Schlotheimer A., Walter S., Schroeder J., Moorman M., Schrempf H.;
RT "The gene encoding the cellulase (Avicelase) Cell from Streptomyces
RL reticuli and analysis of protein domains.";
CC Mol. Microbiol. 5:3611-3621(1992).
CC -1- FUNCTION: THIS ENDOGLUCANASE ACTS ONLY ON CRYSTALLINE CELLULOSE.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC

```

LINKAGES IN CELLULOSE.  
-1- SUBCELLULAR LOCATION: EXISTS IN BOTH MYCELIIUM-ASSOCIATED AND  
EXTRACELLULAR FORMS.  
-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL  
HYDROLASES).  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: X65616; CAA46570.1; -  
PIR: S27985; S27985.  
HSP: P14090; IUP.  
DR InterPro: IPR003305; CBD\_6.  
DR InterPro: IPR001701; Glyco\_hydro\_9.  
DR Pfam: PF02018; CBD\_6; 1.  
DR Pfam: PF00759; Glyco\_hydro\_9; 1.  
DR PROSITE: PS00592; GLYCOSYL\_HYDROL\_F9\_1; FALSE\_NEG.  
DR PROSITE: PS00598; GLYCOSYL\_HYDROL\_F9\_2; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 746 CELLULOSE 1.  
FT DOMAIN 45 170 CELLULOSE-BINDING.  
FT DOMAIN 171 211 LINKER.  
FT DOMAIN 212 746 CATALYTIC.  
FT ACT\_SITE 669 669 BY SIMILARITY.  
FT ACT\_SITE 717 717 BY SIMILARITY.  
FT ACT\_SITE 726 726 BY SIMILARITY.  
SQ SEQUENCE 746 AA; 79999 MW; FDIADBA714A35FD CRC64;

Query Match 19.9%; Score 61.5; DB 1; Length 746;  
Best Local Similarity 30.0%; Pred No. 17;  
Matches 15; Conservative 8; Mismatches 12; Indels 15; Gaps 2;  
OY 7 GGINAWNTITSYDNOICOGKMKCNNTGDPF-----MCPENGSCVDPD 49  
DB 350 GGATWELTSTY-----ERSLTARTGHPALGDTLALPESGNKVPD 391

RESULT 8  
ID CIN2\_RAT STANDARD; PRT; 2005 AA.  
AC P04775;  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE SODIUM CHANNEL PROTEIN, BRAIN II ALPHA SUBUNIT.  
OS SCN2A1 OR SCN2A.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=66146901; Pubmed=3754035;  
RA Noda M., Ikeda T., Kayano T., Suzuki H., Takeshima H., Kurasaki M.,  
Takahashi H., Numa S.;  
"Existence of distinct sodium channel messenger RNAs in rat brain.";   
Nature 320:188-192(1986).  
RL "FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT SODIUM ION  
PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED  
CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE  
MEMBRANE, THE PROTEIN FORMS A SODIUM-SELECTIVE CHANNEL THROUGH  
WHICH NA++ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL  
GRADIENT."  
CC -1- SUBUNIT: THE SODIUM CHANNEL CONSISTS OF A LARGE POLYPEPTIDE AND  
2-3 SMALLER ONES. THIS SEQUENCE REPRESENTS A LARGE POLYPEPTIDE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

-1- DOMAIN: THE SEQUENCE CONTAINS 4 INTERNAL REPEATS, EACH WITH 5  
HYDROPHOBIC SEGMENTS (S1,S2,S3,S5,S6) AND ONE POSITIVELY CHARGED  
SEGMENT (S4). SEGMENTS S4 ARE PROBABLY THE VOLTAGE-SENSORS AND ARE  
CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT  
EVERY THIRD POSITION.  
-1- SIMILARITY: TO OTHER SODIUM CHANNEL PROTEINS.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: X03639; CAA27287.1; -  
PIR: B25019; B25019.  
DR InterPro: IPR002111; Cat\_channel\_TrpL.  
DR InterPro: IPR000636; Cation\_chan\_non\_1lg.  
DR InterPro: IPR001682; Channel\_pore\_ca\_na.  
DR InterPro: IPR00048; IQ.  
DR InterPro: IPR001696; Na\_channel.  
DR Pfam: PF00520; Ion\_trans; 4.  
DR Pfam: PF00612; IQ; 1.  
DR PRINTS: PR00170; NACHANNEL.  
DR SMART: SM00015; IQ; 1.  
DR PROSITE: PS50096; IQ; 1.  
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
KW Glycoprotein; Repeat; Multigene family.  
FT REPEAT 111 456 I.  
FT REPEAT 741 1013 II.  
FT REPEAT 1190 1504 III.  
FT REPEAT 1513 1811 IV.  
FT TRANSMEM 125 148 S1 OF REPEAT I.  
FT TRANSMEM 157 176 S2 OF REPEAT I.  
FT TRANSMEM 190 208 S3 OF REPEAT I.  
FT TRANSMEM 215 234 S4 OF REPEAT I.  
FT TRANSMEM 251 274 S5 OF REPEAT I.  
FT TRANSMEM 402 427 S6 OF REPEAT I.  
FT TRANSMEM 754 778 S1 OF REPEAT II.  
FT TRANSMEM 790 813 S2 OF REPEAT II.  
FT TRANSMEM 822 841 S3 OF REPEAT II.  
FT TRANSMEM 848 867 S4 OF REPEAT II.  
FT TRANSMEM 884 904 S5 OF REPEAT II.  
FT TRANSMEM 958 983 S6 OF REPEAT II.  
FT TRANSMEM 1204 1227 S1 OF REPEAT III.  
FT TRANSMEM 1241 1266 S2 OF REPEAT III.  
FT TRANSMEM 1273 1294 S3 OF REPEAT III.  
FT TRANSMEM 1299 1320 S4 OF REPEAT III.  
FT TRANSMEM 1340 1367 S5 OF REPEAT III.  
FT TRANSMEM 1447 1473 S6 OF REPEAT III.  
FT TRANSMEM 1527 1550 S1 OF REPEAT IV.  
FT TRANSMEM 1562 1585 S2 OF REPEAT IV.  
FT TRANSMEM 1592 1615 S3 OF REPEAT IV.  
FT TRANSMEM 1626 1647 S4 OF REPEAT IV.  
FT TRANSMEM 1663 1685 S5 OF REPEAT IV.  
FT TRANSMEM 1752 1776 S6 OF REPEAT IV.  
FT DOMAIN 1905 1934 IO.  
FT CARBOHYD 212 212 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 291 291 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 297 297 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 303 303 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 308 308 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 340 340 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 604 604 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 624 624 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 883 883 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 1055 1055 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 1072 1072 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 1136 1136 N-LINKED (GLCNAC... ) (POTENTIAL).  
FT CARBOHYD 1368 1368 N-LINKED (GLCNAC... ) (POTENTIAL).



FT	DISULFID	329	343	BY SIMILARITY.
FT	DISULFID	345	358	BY SIMILARITY.
FT	DISULFID	491	503	BY SIMILARITY.
FT	DISULFID	498	512	BY SIMILARITY.
FT	DISULFID	514	526	BY SIMILARITY.
FT	DISULFID	532	542	BY SIMILARITY.
FT	DISULFID	537	551	BY SIMILARITY.
FT	DISULFID	553	566	BY SIMILARITY.
FT	DISULFID	572	584	BY SIMILARITY.
FT	DISULFID	579	593	BY SIMILARITY.
FT	DISULFID	595	608	BY SIMILARITY.
FT	DISULFID	614	625	BY SIMILARITY.
FT	DISULFID	620	634	BY SIMILARITY.
FT	DISULFID	636	649	BY SIMILARITY.
FT	DISULFID	655	666	BY SIMILARITY.
FT	DISULFID	661	675	BY SIMILARITY.
FT	DISULFID	677	690	BY SIMILARITY.
FT	DISULFID	765	777	BY SIMILARITY.
FT	DISULFID	772	786	BY SIMILARITY.
FT	DISULFID	788	801	BY SIMILARITY.
FT	DISULFID	807	819	BY SIMILARITY.
FT	DISULFID	814	828	BY SIMILARITY.
FT	DISULFID	830	843	BY SIMILARITY.
FT	DISULFID	849	859	BY SIMILARITY.
FT	DISULFID	854	868	BY SIMILARITY.
FT	DISULFID	870	883	BY SIMILARITY.
FT	DISULFID	952	964	BY SIMILARITY.
FT	DISULFID	959	973	BY SIMILARITY.
FT	DISULFID	975	988	BY SIMILARITY.
FT	DISULFID	1070	1082	BY SIMILARITY.
FT	DISULFID	1077	1091	BY SIMILARITY.
FT	DISULFID	1093	1106	BY SIMILARITY.
FT	DISULFID	1112	1124	BY SIMILARITY.
FT	DISULFID	1119	1133	BY SIMILARITY.
FT	DISULFID	1135	1149	BY SIMILARITY.
FT	DISULFID	1155	1167	BY SIMILARITY.
FT	DISULFID	1162	1176	BY SIMILARITY.
FT	DISULFID	1178	1191	BY SIMILARITY.
FT	DISULFID	1197	1209	BY SIMILARITY.
FT	DISULFID	1204	1218	BY SIMILARITY.
FT	DISULFID	1220	1233	BY SIMILARITY.
FT	DISULFID	1239	1250	BY SIMILARITY.
FT	DISULFID	1246	1259	BY SIMILARITY.
FT	DISULFID	1261	1274	BY SIMILARITY.
FT	DISULFID	1280	1292	BY SIMILARITY.
FT	DISULFID	1287	1301	BY SIMILARITY.
FT	DISULFID	1303	1316	BY SIMILARITY.
FT	DISULFID	1322	1334	BY SIMILARITY.
FT	DISULFID	1332	1343	BY SIMILARITY.
FT	DISULFID	1345	1358	BY SIMILARITY.
FT	DISULFID	1364	1377	BY SIMILARITY.
FT	DISULFID	1371	1386	BY SIMILARITY.
FT	DISULFID	1388	1399	BY SIMILARITY.
FT	DISULFID	1405	1418	BY SIMILARITY.
FT	DISULFID	1412	1427	BY SIMILARITY.
FT	DISULFID	1429	1440	BY SIMILARITY.
FT	DISULFID	1446	1458	BY SIMILARITY.
FT	DISULFID	1463	1467	BY SIMILARITY.
FT	DISULFID	1469	1482	BY SIMILARITY.
FT	DISULFID	1488	1499	BY SIMILARITY.
FT	DISULFID	1494	1508	BY SIMILARITY.
FT	DISULFID	1510	1523	BY SIMILARITY.
FT	DISULFID	1529	1540	BY SIMILARITY.
FT	DISULFID	1535	1549	BY SIMILARITY.
FT	DISULFID	1551	1564	BY SIMILARITY.
FT	DISULFID	1647	1659	BY SIMILARITY.
FT	DISULFID	1654	1668	BY SIMILARITY.
FT	DISULFID	1670	1683	BY SIMILARITY.
FT	DISULFID	1689	1701	BY SIMILARITY.
FT	DISULFID	1696	1710	BY SIMILARITY.

Query Match 19.9%; Score 61.5; DB 1; Length 2907;  
 Best Local Similarity 40.4%; Pred. No. 59;

Matches 19: Conservative 3; Mismatches 10; Indels 15; Gaps 4;  
 QY 19 IDNQCQCG-----KMLCUNT-----GDE-EMCPENGSCVPPGPG 52  
 Db 220 VNMOCQGLTIGVCTKTLCCATIGRANGHPCMCAPAPQ--PCRPQ 264

RESULT 10  
 LRP2\_HUMAN STANDARD; PRT: 4655 AA.  
 ID LRP2\_HUMAN  
 AC P98164; 000711; Q16215;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED PROTEIN 2 PRECURSOR (MEGALIN)  
 DE (GLYCOPROTEIN 330).  
 GN LRP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Kidney;  
 RX MEDLINE=96305376; PubMed=8706697;  
 RA Hjaelén G., Murray E., Crumley G., Harazin W., Lundgren S., Onyango I.,  
 RA Ek B., Larsson M., Juhlin C., Hellman P., Davis H., Akerstrom G.,  
 RA Rask L., Morse B.;  
 RT "Cloning and sequencing of human gp330, a Ca(2+)-binding receptor  
 RT with potential intracellular signaling properties.";  
 RL Eur. J. Biochem. 239:132-137(1996).  
 RN [2]  
 RP SEQUENCE OF 2705-4453 FROM N.A.  
 RC TISSUE-Kidney;  
 RX MEDLINE=95048397; PubMed=7959795;  
 RA Koraak C., Argaves W.S.;  
 RL Submitted (DEC-1993) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 3833-4453 FROM N.A.  
 RC TISSUE-Kidney;  
 RX MEDLINE=95048397; PubMed=7959795;  
 RA Koraak C., Argaves W.S.;  
 RL Submitted (DEC-1993) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 4139-4406 FROM N.A.  
 RX MEDLINE=94244704; PubMed=8187828;  
 RA Lundgren S., Hjaelén G., Hellman P., Ek B., Juhlin C., Rastad J.,  
 RA Klarskog L., Akerstrom G., Rask L.;  
 RT "A protein involved in calcium sensing of the human parathyroid and  
 RT placental cytotrophoblast cells belongs to the LDL-receptor protein  
 RT superfamily.";  
 RL Genomics 22:88-93(1994).  
 RN [5]  
 RP EXP. Cell Res. 212:344-350(1994).  
 RX MEDLINE=95286588; PubMed=7768901;  
 RA Kounnas M.Z., Strickland D.K., Argaves W.S.;  
 RA Brewer B.H., Strickland D.K., Argaves W.S.;  
 RT "Identification of glycoprotein 330 as an endocytic receptor for  
 RT apolipoprotein J/clusterin.";  
 RL J. Biol. Chem. 270:13070-13075(1995).  
 CC -1- FUNCTION: BINDS SPECIFICALLY CLUSTERIN WITH HIGH AFFINITY, BUT  
 CC ALSO LIGANDS IN COMMON WITH OTHER FAMILY MEMBERS: PLASMINOGEN,  
 CC EXTRACELLULAR MATRIX COMPONENTS, PLASMINOGEN ACTIVATOR-PLASMINOGEN  
 CC ACTIVATOR INHIBITOR TYPE 1 COMPLEX, APOLOPROTEIN E-ENRICHED  
 CC BETA-VLDL, LIPOPROTEIN LIPASE, LACTOFERRIN AND CALCIUM.  
 CC -1- FUNCTION: RECEPTOR-MEDIATED UPTAKE OF POLYBASIC DRUGS SUCH AS  
 CC APROTININ, AMINOGLYCOSIDES AND POLYMYXIN B (BY SIMILARITY).  
 CC -1- FUNCTION: MAY PARTICIPATE IN REGULATION OF PARATHYROID-HORMONE AND  
 CC PARA-THYROID-HORMONE-RELATED PROTEIN RELEASE.  
 CC -1- SUBUNIT: FORMS A MULTIMERIC COMPLEX TOGETHER WITH A RECEPTOR-

CC ASSOCIATED PROTEIN (RAP).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: ABSORPTIVE EPITHELIA, INCLUDING RENAL  
 CC PROXIMAL TUBULES.  
 CC -1- SIMILARITY: CONTAINS 36 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 37 LDL-RECEPTOR CLASS B DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 17 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 4 SH3-BINDING DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 SH2-BINDING DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC -----  
 DR EMBL: U03837; ABA41649.1; -;  
 DR EMBL: U04441; AAB02882.1; -;  
 DR EMBL: S73145; AAB30825.1; -;  
 DR HSP: P01130; IAUJ.  
 DR MIM: 600073; -;  
 DR Interpro: IPR000152; Asx\_hydroxyl.  
 DR Interpro: IPR000561; EGF-like.  
 DR Interpro: IPR001881; EGF-Ca.  
 DR Interpro: IPR002172; LDL\_recept\_A.  
 DR Interpro: IPR000033; Ldl\_receptor\_rep.  
 DR Pfam: PF00008; EGF\_13.  
 DR Pfam: PF00057; Ldl\_recept\_a; 36.  
 DR Pfam: PF00058; Ldl\_recept\_b; 37.  
 DR PRINTS: PR00261; LDLRECEPTOR.  
 DR SMART: SM00179; EGF\_CA; 3.  
 DR SMART: SM00001; EGF\_Like; 15.  
 DR SMART: SM00192; LDLr; 36.  
 DR PROSITE: PS00135; LY; 35.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 4.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS01186; EGF\_2; 9.  
 DR PROSITE: PS01187; EGF\_CA; 3.  
 DR PROSITE: PS01209; LDLR\_1; 31.  
 DR PROSITE: PS50068; LDLR\_2; 36.  
 DR GlycoProtein: Repeat; Endocytosis; Coated pits; Transmembrane;  
 KW Receptor; EGF-like domain; Signal; Polymorphism.  
 FT SIGNAL 1 25  
 FT CHAIN 26 4655  
 FT FT 26 4655 LOW-DENSITY LIPOPROTEIN RECEPTOR-RELATED  
 FT FT 26 4655 PROTEIN 2.  
 FT FT 26 4655 EXTRACELLULAR (POTENTIAL).  
 FT FT 26 4655 POTENTIAL.  
 FT FT 26 4655 CYTOPLASMIC (POTENTIAL).  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 1.  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 2.  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 3.  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 4.  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 5.  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 6.  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 7.  
 FT FT 26 4655 EGF-Like 1.  
 FT FT 26 4655 EGF-Like 2.  
 FT FT 26 4655 LDL-RECEPTOR CLASS B 1.  
 FT FT 26 4655 LDL-RECEPTOR CLASS B 2.  
 FT FT 26 4655 LDL-RECEPTOR CLASS B 3.  
 FT FT 26 4655 LDL-RECEPTOR CLASS B 4.  
 FT FT 26 4655 LDL-RECEPTOR CLASS B 5.  
 FT FT 26 4655 EGF-Like 3.  
 FT FT 26 4655 LDL-RECEPTOR CLASS B 6.  
 FT FT 26 4655 LDL-RECEPTOR CLASS B 7.  
 FT FT 26 4655 LDL-RECEPTOR CLASS B 8.  
 FT FT 26 4655 LDL-RECEPTOR CLASS B 9.  
 FT FT 26 4655 EGF-Like 4.  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 8.  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 9.  
 FT FT 26 4655 LDL-RECEPTOR CLASS A 10.

FT DOMAIN 1147 1185 LDL-RECEPTOR CLASS A 11.  
 FT DOMAIN 1186 1224 LDL-RECEPTOR CLASS A 12.  
 FT DOMAIN 1228 1268 LDL-RECEPTOR CLASS A 13.  
 FT DOMAIN 1269 1307 LDL-RECEPTOR CLASS A 14.  
 FT DOMAIN 1310 1350 LDL-RECEPTOR CLASS A 15.  
 FT DOMAIN 1349 1389 EGF-Like 5.  
 FT DOMAIN 1350 1429 EGF-Like 6.  
 FT DOMAIN 1478 1519 LDL-RECEPTOR CLASS B 10.  
 FT DOMAIN 1521 1562 LDL-RECEPTOR CLASS B 11.  
 FT DOMAIN 1608 1653 LDL-RECEPTOR CLASS B 12.  
 FT DOMAIN 1655 1695 LDL-RECEPTOR CLASS B 13.  
 FT DOMAIN 1700 1741 EGF-Like 7.  
 FT DOMAIN 1790 1831 LDL-RECEPTOR CLASS B 15.  
 FT DOMAIN 1833 1881 LDL-RECEPTOR CLASS B 16.  
 FT DOMAIN 1883 1929 LDL-RECEPTOR CLASS B 17.  
 FT DOMAIN 1931 1971 LDL-RECEPTOR CLASS B 18.  
 FT DOMAIN 1972 2012 LDL-RECEPTOR CLASS B 19.  
 FT DOMAIN 2018 2059 EGF-Like 8.  
 FT DOMAIN 2107 2155 LDL-RECEPTOR CLASS B 20.  
 FT DOMAIN 2157 2200 LDL-RECEPTOR CLASS B 21.  
 FT DOMAIN 2202 2244 LDL-RECEPTOR CLASS B 22.  
 FT DOMAIN 2246 2288 LDL-RECEPTOR CLASS B 23.  
 FT DOMAIN 2290 2331 LDL-RECEPTOR CLASS B 24.  
 FT DOMAIN 2342 2383 EGF-Like 9.  
 FT DOMAIN 2431 2476 LDL-RECEPTOR CLASS B 25.  
 FT DOMAIN 2478 2517 LDL-RECEPTOR CLASS B 26.  
 FT DOMAIN 2519 2561 LDL-RECEPTOR CLASS B 27.  
 FT DOMAIN 2563 2603 LDL-RECEPTOR CLASS B 28.  
 FT DOMAIN 2604 2646 LDL-RECEPTOR CLASS B 29.  
 FT DOMAIN 2651 2693 EGF-Like 10.  
 FT DOMAIN 2698 2738 LDL-RECEPTOR CLASS A 16.  
 FT DOMAIN 2739 2777 LDL-RECEPTOR CLASS A 17.  
 FT DOMAIN 2778 2819 LDL-RECEPTOR CLASS A 18.  
 FT DOMAIN 2820 2861 LDL-RECEPTOR CLASS A 19.  
 FT DOMAIN 2862 2901 LDL-RECEPTOR CLASS A 20.  
 FT DOMAIN 2904 2945 LDL-RECEPTOR CLASS A 21.  
 FT DOMAIN 2946 2990 LDL-RECEPTOR CLASS A 22.  
 FT DOMAIN 2991 3029 LDL-RECEPTOR CLASS A 23.  
 FT DOMAIN 3030 3070 LDL-RECEPTOR CLASS A 24.  
 FT DOMAIN 3073 3110 LDL-RECEPTOR CLASS A 25.  
 FT DOMAIN 3111 3151 EGF-Like 11.  
 FT DOMAIN 3152 3192 EGF-Like 12.  
 FT DOMAIN 3239 3280 LDL-RECEPTOR CLASS B 30.  
 FT DOMAIN 3282 3331 LDL-RECEPTOR CLASS B 31.  
 FT DOMAIN 3333 3375 LDL-RECEPTOR CLASS B 32.  
 FT DOMAIN 3377 3418 LDL-RECEPTOR CLASS B 33.  
 FT DOMAIN 3419 3459 LDL-RECEPTOR CLASS B 34.  
 FT DOMAIN 3465 3509 EGF-Like 13.  
 FT DOMAIN 3510 3550 LDL-RECEPTOR CLASS A 26.  
 FT DOMAIN 3551 3591 LDL-RECEPTOR CLASS A 27.  
 FT DOMAIN 3592 3632 LDL-RECEPTOR CLASS A 28.  
 FT DOMAIN 3633 3673 LDL-RECEPTOR CLASS A 29.  
 FT DOMAIN 3676 3716 LDL-RECEPTOR CLASS A 30.  
 FT DOMAIN 3717 3756 LDL-RECEPTOR CLASS A 31.  
 FT DOMAIN 3757 3795 LDL-RECEPTOR CLASS A 32.  
 FT DOMAIN 3796 3834 LDL-RECEPTOR CLASS A 33.  
 FT DOMAIN 3840 3880 LDL-RECEPTOR CLASS A 34.  
 FT DOMAIN 3881 3922 LDL-RECEPTOR CLASS A 35.  
 FT DOMAIN 3926 3964 LDL-RECEPTOR CLASS A 36.  
 FT DOMAIN 3966 4006 EGF-Like 14.  
 FT DOMAIN 4007 4048 EGF-Like 15.  
 FT DOMAIN 4154 4195 CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 4195 LDL-RECEPTOR CLASS B 35.

Query Match 19.7%; Score 61; DB 1; Length 4655;  
 Best Local Similarity 23.0%; Pred. No. 1e+02;  
 Matches 17; Conservative 7; Mismatches 16; Indels 34; Gaps 4;  
 QY 5 CPGGINAMNTITSYINDIOGQGNLCNNMGPE-----MCPENGSCV 47  
 Db 229 CPFG-----RITYNWNWCDGEDD-CRDNQDCEGSGPHVHKCSPRENSCPESGKCI 280  
 QY 48 P-----DGGP 52

DB 281 SIYKWDGILDCPG 294

RESULT 11  
CIN2\_HUMAN STANDARD; PRT; 2005 AA.

AC 099250; Q1472;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE SODIUM CHANNEL PROTEIN, BRAIN II ALPHA SUBUNIT.  
GN SCN2A1 OR SCN2A OR NAC2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
MEDLINE=92390418; PubMed=1325650;  
RA Ahmed C.M., Ware D.H., Lee S.C., Patten C.D., Ferrer-Montiel A.V.,  
RA Schindler A.F., McPherson J.D., Wagner-McPherson C.B., Wasmuth J.J.,  
RA Evans G.A., Montal M.;  
RT "Primary structure, chromosomal localization, and functional  
RT expression of a voltage-gated sodium channel from human brain.";  
RT Proc. Natl. Acad. Sci. U.S.A. 89:8220-8224(1992).  
RN [2]  
RP SEQUENCE OF 1-89 FROM N.A.  
RA Lu C.-M., Eichelberger J.S., Beckman M.L., Schade S.D., Brown G.B.;  
RT "Isolation of the 5'-flanking region for human brain sodium channel  
RT subtype II alpha-subunit (SCN2A).";  
RT Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1702-2005 FROM N.A.  
RC TISSUE=Brain;  
MEDLINE=92275082; PubMed=1317301;  
RA Lu C.-M., Han J., Rado T.A., Brown G.B.;  
RT "Differential expression of two sodium channel subtypes in human  
RT brain.";  
RT FEBS Lett. 303:53-58(1992).  
RN [4]  
RP SEQUENCE OF 1702-1772 FROM N.A.  
MEDLINE=91110524; PubMed=1846440;  
RA Han J., Lu C.-M., Brown G.B., Rado T.A.;  
RT "Direct amplification of a single dissected chromosomal segment by  
RT polymerase chain reaction: a human brain sodium channel gene is on  
RT chromosome 2q22-q23.";  
RT Proc. Natl. Acad. Sci. U.S.A. 88:335-339(1991).  
RN [5]  
RP FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT SODIUM ION  
PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED  
CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE  
MEMBRANE, THE PROTEIN FORMS A SODIUM-SELECTIVE CHANNEL THROUGH  
WHICH NA++ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL  
GRADIENT.  
CC -1- SUBUNIT: THE SODIUM CHANNEL CONSISTS OF A LARGE POLYPEPTIDE AND  
CC 2-3 SMALLER ONES. THIS SEQUENCE REPRESENTS A LARGE POLYPEPTIDE.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -1- DOMAIN: THE SEQUENCE CONTAINS 4 INTERNAL REPEATS, EACH WITH 5  
CC HYDROPHOBIC SEGMENTS (S1,S2,S3,S5,S6) AND ONE POSITIVELY CHARGED  
CC SEGMENT (S4). SEGMENTS S4 ARE PROBABLY THE VOLTAGE-SENSORS AND ARE  
CC CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT  
CC EVERY THIRD POSITION.  
CC -1- SIMILARITY: TO OTHER SODIUM CHANNEL PROTEINS.  
CC -1- SIMILARITY: CONTAINS 1 TO DOMAIN.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
DR EMBL; M94055; AAA18895.1; -  
DR EMBL; AF059683; AAC14574.1; -  
DR EMBL; X65361; CAA46438.1; ALT\_SRO.  
DR EMBL; M91804; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; M55662; AAB65854.2; -  
DR PIR; S29185; S29185.  
DR MIM; 182390; -  
DR InterPro; IPR002111; Cat\_channel\_TrypL.  
DR InterPro; IPR000636; Cation\_chan\_non\_lig.  
DR InterPro; IPR001682; Channel\_pore\_Ca\_Na.  
DR InterPro; IPR000048; IQ.  
DR InterPro; IPR001696; Na\_channel.  
DR Pfam; PF00520; Ion\_trans; 4.  
DR Pfam; PF00612; IQ; 2.  
DR PRINTS; PR00170; NACHANNEL.  
DR SMART; SM00015; IQ; 1.  
DR PROSITE; PS0096; IQ; 1.  
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
KW Glycoprotein; Repeat; Multigene family.  
FT REPEAT 111 456  
FT REPEAT 741 1013  
FT REPEAT 1190 1504  
FT REPEAT 1513 1811  
FT REPEAT 125 148  
FT TRANSMEM 157 176  
FT TRANSMEM 190 208  
FT TRANSMEM 215 234  
FT TRANSMEM 251 274  
FT TRANSMEM 402 427  
FT TRANSMEM 754 778  
FT TRANSMEM 790 813  
FT TRANSMEM 822 841  
FT TRANSMEM 848 867  
FT TRANSMEM 884 904  
FT TRANSMEM 958 983  
FT TRANSMEM 1204 1227  
FT TRANSMEM 1241 1266  
FT TRANSMEM 1273 1294  
FT TRANSMEM 1299 1320  
FT TRANSMEM 1340 1367  
FT TRANSMEM 1447 1473  
FT TRANSMEM 1527 1550  
FT TRANSMEM 1562 1585  
FT TRANSMEM 1592 1615  
FT TRANSMEM 1626 1647  
FT TRANSMEM 1663 1685  
FT TRANSMEM 1752 1776  
FT DOMAIN 1905 1934  
FT CARBOHYD 212 212  
FT CARBOHYD 285 285  
FT CARBOHYD 291 291  
FT CARBOHYD 297 297  
FT CARBOHYD 303 303  
FT CARBOHYD 308 308  
FT CARBOHYD 340 340  
FT CARBOHYD 604 604  
FT CARBOHYD 624 624  
FT CARBOHYD 883 883  
FT CARBOHYD 1055 1055  
FT CARBOHYD 1072 1072  
FT CARBOHYD 1136 1136  
FT CARBOHYD 1368 1368  
FT CARBOHYD 1382 1382  
FT CARBOHYD 1393 1393  
FT CARBOHYD 1778 1778  
FT CARBOHYD 1768 1768  
FT CONFLICT 1990 1990  
FT CONFLICT 1995 1995  
SQ SEQUENCE 2005 AA; 227916 MW; 844BBE88A4FA21 CRC64;  
19.6%; Score 60.5; DB 1; Length 2005;





```

FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 460 474 BY SIMILARITY.
FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 776 790 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.

```

```

Query Match 19.6%; Score 60.5; DB 1; Length 2871;
Best Local Similarity 47.2%; Pred. NO. 75;
Matches 17; Conservative 1; Mismatches 5; Indels 13; Gaps 3;

```

```

QY 19 IDNOCGCG-----KNLCNNT-----GDP-EMCP 41
DB 190 ISNOCGCGSLGIVCTKTLCCATVGRAMHPCMCMP 225

```

```

RESULT 13
FBN2_HUMAN STANDARD; PRT; 2911 AA.
AC P35556;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FIBRILLIN 2 PRECURSOR.
GN FBN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165150; PubMed=8120105;
RA Zhang H., Apfelroth S.D., Hu W., Davis E.C., Sanguineti C.,
RA Bonadio J., Mecham R.P., Ramirez F.;
RT "Structure and expression of fibrillin-2, a novel microfibrillar
RT component preferentially located in elastic matrices.";
RL J. Cell Biol. 124:855-863(1994).
RN [2]
RP SEQUENCE OF 752-1505 FROM N.A.
RX MEDLINE=91304567; PubMed=1852206;
RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
RA Tsipouras P., Ramirez F., Hollister D.;
RT "Linkage of Marfan syndrome and a phenotypically related disorder to
RT two different fibrillin genes.";
RL Nature 352:330-334(1991).
RN [3]
RP VARIANTS CCA TYR-1252 AND SER-1433, AND VARIANT ILE-964.
RX MEDLINE=96083599; PubMed=7493032;
RA Putnam E.A., Zhang H., Ramirez F., Mlewicz D.M.;
RT "Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder,
RT congenital contractural arachnodactyly.";
RL Nat. Genet. 11:456-458(1995).
RN [4]
RP VARIANTS CCA HIS-1114.
RX MEDLINE=98407789; PubMed=9737771;
RA Babcock D., Gasner C., Francke U., Maslen C.;
RT "A single mutation that results in an asp-to-his substitution and
RT partial exon skipping in a family with congenital contractural
RT arachnodactyly.";
RL Hum. Genet. 103:22-28(1998).
RN [5]
RP VARIANTS CCA PHE-1141 AND TRP-1252.
RX MEDLINE=20259236; PubMed=10797416;
RA Bellet S., Zhou G., Wang M., Der Kaloustian V.M., Pagon R.A.,
RA Godfrey M.;
RT "Two novel fibrillin-2 mutations in congenital contractural
RT arachnodactyly.";
RL Am. J. Med. Genet. 92:7-12(2000).
RN [6]
RP FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
RN THAT BINDS CALCIUM. FIBRILLIN-2-CONTAINING MICROFIBRILS REGULATE
RN THE EARLY PROCESS OF ELASTIC FIBER ASSEMBLY.
RN [7]
RP DISEASE: DEFECTS IN FBN2 ARE THE CAUSE OF CONGENITAL CONTRACTURAL
RN ARACHNOACTYLY (CCA) (ALSO KNOWN AS BEALS SYNDROME). CCA IS
RN PHENOTYPICALLY SIMILAR TO MARFAN SYNDROME, BUT DOES NOT AFFECT THE
RN AORTA AND THE EYES.
RN [8]
RP SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
RN EGF-LIKE DOMAINS.
RN [9]
RP SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
RN [10]
RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (see http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [11]
RP EMBL: U03272; AAA18950.1;

```



DR EMBL: X62009; -, NOT\_ANNOTATED\_CDS.  
 DR PIR: S17063; S17063.  
 DR PIR: S31101; S31101.  
 DR HSSP: P35555; 1EMO.  
 DR MIM: 121050; -.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR001438; EGF-11.  
 DR InterPro: IPR002212; TB.  
 DR Pfam: PF00008; EGF; 46.  
 DR Pfam: PF00683; TB; 9.  
 DR PRINTS: PRO0010; EGFBL00.  
 DR SMART: SM00179; EGF\_CA; 43.  
 DR SMART: SM00001; EGF\_Like; 3.  
 DR PROSITE: PS00020; ASX\_HYDROXYL; 43.  
 DR PROSITE: PS00022; EGF\_1; 2.  
 DR PROSITE: PS01186; EGF\_2; 37.  
 DR PROSITE: PS01187; EGF\_CA; 43.  
 DR Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;.  
 KW Repeat; Signal; Multigene family; Disease mutation; Polymorphism.  
 FT SIGNAL 1 28  
 FT CHAIN 1 29 2911  
 FT DOMAIN 111 142  
 FT DOMAIN 145 176  
 FT DOMAIN 176 207  
 FT DOMAIN 275 316  
 FT DOMAIN 317 358  
 FT REPEAT 359 425  
 FT DOMAIN 493 533  
 FT DOMAIN 534 573  
 FT DOMAIN 574 615  
 FT DOMAIN 616 656  
 FT DOMAIN 657 697  
 FT REPEAT 698 766  
 FT DOMAIN 767 808  
 FT DOMAIN 809 850  
 FT DOMAIN 851 890  
 FT DOMAIN 954 995  
 FT REPEAT 996 1071  
 FT DOMAIN 1072 1113  
 FT DOMAIN 1114 1156  
 FT DOMAIN 1157 1198  
 FT DOMAIN 1199 1240  
 FT DOMAIN 1241 1281  
 FT DOMAIN 1282 1323  
 FT DOMAIN 1324 1365  
 FT DOMAIN 1366 1406  
 FT DOMAIN 1407 1447  
 FT DOMAIN 1448 1489  
 FT DOMAIN 1490 1530  
 FT DOMAIN 1531 1571  
 FT REPEAT 1572 1648  
 FT DOMAIN 1649 1690  
 FT DOMAIN 1691 1732  
 FT REPEAT 1733 1806  
 FT DOMAIN 1807 1848  
 FT DOMAIN 1849 1890  
 FT DOMAIN 1891 1932  
 FT DOMAIN 1933 1971  
 FT DOMAIN 1972 2014  
 FT DOMAIN 2015 2054  
 FT DOMAIN 2055 2096  
 FT REPEAT 2097 2169  
 FT DOMAIN 2170 2211  
 FT DOMAIN 2212 2251  
 FT DOMAIN 2252 2292  
 FT DOMAIN 2293 2336  
 FT DOMAIN 2337 2378  
 FT REPEAT 2379 2447  
 FT DOMAIN 2448 2489  
 FT DOMAIN 2490 2530  
 FT DOMAIN 2531 2569

FT DOMAIN 2570 2612  
 FT DOMAIN 2613 2652  
 FT DOMAIN 2653 2693  
 FT DOMAIN 2694 2733  
 FT DISULFID 115 124  
 FT DISULFID 119 130  
 FT DISULFID 132 141  
 FT DISULFID 149 159  
 FT DISULFID 153 164  
 FT DISULFID 166 175  
 FT DISULFID 180 190  
 FT DISULFID 184 195  
 FT DISULFID 197 206  
 FT DISULFID 279 291  
 FT DISULFID 286 300  
 FT DISULFID 302 315  
 FT DISULFID 321 333  
 FT DISULFID 328 342  
 FT DISULFID 344 357  
 FT DISULFID 497 509  
 FT DISULFID 504 518  
 FT DISULFID 520 532  
 FT DISULFID 538 548  
 FT DISULFID 543 557  
 FT DISULFID 559 572  
 FT DISULFID 578 585  
 FT DISULFID 585 599  
 FT DISULFID 601 614  
 FT DISULFID 620 631  
 FT DISULFID 626 640  
 FT DISULFID 642 655  
 FT DISULFID 661 672  
 FT DISULFID 667 681  
 FT DISULFID 683 696  
 FT DISULFID 771 783  
 FT DISULFID 778 792  
 FT DISULFID 794 807  
 FT DISULFID 813 825  
 FT DISULFID 820 834  
 FT DISULFID 836 849  
 FT DISULFID 855 865  
 FT DISULFID 860 874  
 FT DISULFID 876 889  
 FT DISULFID 958 970  
 FT DISULFID 965 979  
 FT DISULFID 981 994  
 FT DISULFID 1076 1088  
 FT DISULFID 1083 1097  
 FT DISULFID 1099 1112  
 FT DISULFID 1118 1130  
 FT DISULFID 1125 1139  
 FT DISULFID 1141 1155  
 FT DISULFID 1161 1173  
 FT DISULFID 1168 1182  
 FT DISULFID 1184 1197  
 FT DISULFID 1203 1215  
 FT DISULFID 1210 1224

Query Match Best Local Similarity 44.48; Score 60.5; DB 1; Length 2911;  
 Pred. No. 76; Mismatches 16; Conservative 3; Indels 13; Gaps 3;

QY 19 IDNQCQGG-----KNLCNNT-----GDP-EMCP 41  
 Db 219 VNNQMGQGLTGIVCTKTLCCATTCGRAGHPCMC 254

RESULT 14  
 ID TX26\_PHONI STANDARD; PRT: 48 AA.  
 AC P29425;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-APR-1993 (Rel. 25, last annotation update)  
 DE NEUROTOXIN TX2-6.  
 CC Phenocytia nigriverter (Brazilian armed spider).  
 CC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 CC Araneomorphae; Entelegynae; Lycosidae; Ctenidae; Phoneytria.  
 CC NCBI\_TaxID=6918;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-Venom;  
 RX MEDLINE=93011905; PubMed=1397265;  
 RA Cordeiro M.N., Diniz C.R., Valentim A.C., von Eickstedt V.R.D.,  
 RA Gilroy J., Richardson M.;  
 RT "The purification and amino acid sequences of four Tx2 neurotoxins  
 from the venom of the Brazilian 'armed' spider Phoneytria nigriverter  
 (Keys)";  
 RT FEBS Lett. 310:153-156(1992).  
 CC -1- FUNCTION: CAUSES SCRATCHING, LACRIMATION, HYPERSALIVATION,  
 CC SWEATING AND AGITATION FOLLOWED BY SPASTIC PARALYSIS OF THE  
 CC ANTERIOR AND POSTERIOR EXTREMITIES AND DEATH AT DOSE LEVELS OF  
 CC 0.79 MG/MOUSE. IT SIGNIFICANTLY ACTIVATES VOLTAGE-DEPENDENT SODIUM  
 CC CHANNELS. INSECTICIDAL TO THE LARVAL AND ADULT FORMS OF THE HOUSE  
 CC FLY.  
 CC -1- SIMILARITY: TO NEUROTOXINS TX2-1 AND TX2-5.  
 DR PIR; S29216; S29216.  
 KW Venom; Neurotoxin.  
 SQ SEQUENCE 48 AA; 5298 MW; 4F30EBFCE4919CD CRC64;

Query Match 19.4%; Score 60; DB 1; Length 48;  
 Best Local Similarity 46.4%; Pred. No. 1.9; Mismatches 13; Indels 2; Gaps 1;  
 Matches 13; Conservative 0;

DT 24 COGOKNLNNTGDPKPCPENGSCVDPGP 51  
 Db 3 CAGDDQPKCTCD--CCGERGECVCGCP 28

RESULT 15  
 ID COAT\_SBMV STANDARD; PRT; 279 AA.  
 AC P03607;  
 DT 21-JUL-1986 (Rel. 01, Created).  
 DT 01-MAY-1991 (Rel. 18, last sequence update)  
 DT 15-JUL-1999 (Rel. 38, last annotation update)  
 DE COAT PROTEIN PRECURSOR (CAPSID PROTEIN).  
 OS Southern bean mosaic virus (SBMV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Sobemovirus.  
 OX NCBI\_TaxID=12139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COMPEA;  
 RX MEDLINE=88044510; PubMed=2823471;  
 RA Wu S., Rinehart C.A., Kaesberg P.;  
 RT "Sequence and organization of southern bean mosaic virus genomic  
 RNA";  
 RT Virology 161:73-80(1987).  
 RN [2]  
 RP SEQUENCE OF 20-279.  
 RA Hermosino M.A., Abad-Zapatero C., Abdel-Meguid S.S., Pundak S.,  
 RA Rossmann M.G., Tremaine J.H.;  
 RT "Amino acid sequence of southern bean mosaic virus coat protein and  
 its relation to the three-dimensional structure of the virus";  
 RT Virology 119:133-149(1982).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RA Abad-Zapatero C., Abdel-Meguid S.S., Johnson J.E., Leslie A.G.W.,  
 RA Rayment I., Rossmann M.G., Suck D., Tsukihara T.;  
 RT "Structure of southern bean mosaic virus at 2.8-A resolution.";  
 RT Nature 286:33-39(1980).  
 CC -1- SUBUNIT: THE VIRUS PARTICLE HAS A SHELL OF 180 PROTEIN SUBUNITS  
 CC ARRANGED WITH T=3 ICOSAEDRAL SYMMETRY AROUND A CORE OF RNA. EACH  
 CC ICOSAEDRAL UNIT CONTAINS THREE PROTEIN SUBUNITS.  
 CC -1- SIMILARITY: COAT PROTEINS FROM CARMOVIRUSES (CARMV, MNSV, TCV),

CC DIANTHOVIRUSES (RCMV), SOBEMOVIRUSES (MCMV, SBMV), TOMBUVIRUSES  
 CC (AMCV, CNV, CYRSV, AND TBSV), AND TTV ARE RELATED IN THEIR S  
 CC REGION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M23021; AAA46567.1;  
 DR PIR; A04218; VCBW.  
 DR PIR; D33739; VCBWSC.  
 DR PDB; 4SBV; 19-APR-89.  
 DR InterPro: IPR000937; Viral\_coat.  
 DR Pfam: PF00729; Viral\_coat.1.  
 DR PRINTS: PR00233; ICOSAEDRAL.  
 DR PROSITE: PS00555; ICOSA VIR COAT S; 1.  
 KW Coat protein; 3D-structure; Acetylation.  
 FT PROPEP 1  
 FT CHAIN 19  
 FT MOD\_RES 20 279  
 FT MOD\_RES 20 20  
 FT DOMAIN 20 83  
 FT DOMAIN 84 269  
 FT DOMAIN 270 279  
 FT VARIANT 101 101  
 FT VARIANT 245 245  
 FT STRAND 78 81  
 FT STRAND 86 99  
 FT STRAND 105 109  
 FT STRAND 112 115  
 FT HELIX 117 123  
 FT HELIX 124 125  
 FT STRAND 126 126  
 FT STRAND 144 140  
 FT STRAND 148 145  
 FT STRAND 154 155  
 FT STRAND 158 159  
 FT TURN 166 169  
 FT HELIX 170 171  
 FT TURN 173 174  
 FT TURN 176 179  
 FT STRAND 180 181  
 FT TURN 184 186  
 FT TURN 187 191  
 FT HELIX 199 200  
 FT STRAND 203 205  
 FT STRAND 208 210  
 FT STRAND 216 217  
 FT STRAND 221 230  
 FT HELIX 232 234  
 FT TURN 235 238  
 FT HELIX 242 247  
 FT STRAND 242 247  
 FT STRAND 256 269  
 FT STRAND 272 272  
 FT STRAND 275 277  
 FT HELIX 277 277  
 SQ SEQUENCE 279 AA; 30527 MW; 9B1BD427642493C1 CRC64;

Query Match 19.4%; Score 60; DB 1; Length 279;  
 Best Local Similarity 27.3%; Pred. No. 9.8;  
 Matches 12; Conservative 8; Mismatches 18; Indels 6; Gaps 2;

DT 6 PGGINANTTSTYDNCICOGKNLN--NNTGDPKPCPENGSCV 47  
 Db 163 PVSVALSLNKGVTGPGWEGSGLCFVNNT---KCPPTSRAI 202

• Thu Mar 28 09:21:39 2002

us-09-726-348-2\_copy\_126\_177\_1.rsp

Job time: 882 sec

...

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:25:36 ; Search time 21.01 Seconds  
(without alignments)  
90.746 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_126\_177  
Perfect score: 52  
Sequence: 1 QHVNCPGGINAMNTITSYID.....NTGDPKPCPENSVCVPDGP 52

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 3664827 residues  
Word size: 0

Total number of hits satisfying chosen parameters: 1802

Minimum DB seq length: 0  
Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database: SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4	7.7	10	1	TKSL_AEDAE
2	4	7.7	18	1	TKN2_SCYCA
3	4	7.7	19	1	THA_PERAM
4	4	7.7	20	1	MIF_PIG
5	4	7.7	21	1	EPAB_HUMAN
6	4	7.7	25	1	PA22_BOTSC
7	4	7.7	25	1	UBL1_BOVIN
8	4	7.7	29	1	AL21_HORSE
9	3	5.8	7	1	FAR1_HELT
10	3	5.8	8	1	UF06_MOUSE
11	3	5.8	9	1	CONO_CONGE
12	3	5.8	9	1	CONO_CONST
13	3	5.8	9	1	DNFL_LOEMI
14	3	5.8	9	1	ISOT_CYPCA
15	3	5.8	9	1	OXYA_SCYCA
16	3	5.8	9	1	OXYA_SQUAC
17	3	5.8	9	1	OXYF_SCYCA
18	3	5.8	9	1	OXYT_CYPCA
19	3	5.8	9	1	OXYT_EISEO
20	3	5.8	9	1	OXYT_OCTYU
21	3	5.8	9	1	OXYT_RABIT
22	3	5.8	9	1	OXYV_RATCL
23	3	5.8	9	1	OXYV_SQUAC
24	3	5.8	9	1	SAP_STOYA
25	3	5.8	10	1	COXO_RAT
26	3	5.8	10	1	COXO_THIOB
27	3	5.8	10	1	PPCK_FASHE
28	3	5.8	10	1	RPLI_PHOXY
29	3	5.8	10	1	TKS2_AEDAE
30	3	5.8	10	1	UHA3_HUMAN
31	3	5.8	11	1	MORN_HUMAN
32	3	5.8	12	1	UKA2_HUMAN
33	3	5.8	13	1	LIGA_TRAVE

34	3	5.8	13	1	LIGA_TRAVE	P20012	trameles ve
35	3	5.8	13	1	NEUT_CHICK	P13724	gallus gall
36	3	5.8	13	1	NEUT_TRIUV	P31745	trichosurus
37	3	5.8	13	1	TAI3_TREME	P01370	tremella me
38	3	5.8	13	1	UNI2_CLOPA	P81353	clostridium
39	3	5.8	14	1	CAT2_FASHE	P80342	fasciola he
40	3	5.8	14	1	GLG5_SPIOL	P55235	spiniacia ol
41	3	5.8	14	1	TAT_HV1W2	P12509	human immun
42	3	5.8	14	1	TAT_HV1W2	P12511	human immun
43	3	5.8	14	1	UC18_MAIZE	P80624	zea mays (m
44	3	5.8	15	1	FKB7_PIMPS	P81104	pinus pinas
45	3	5.8	15	1	KIOW_LMWTE	P11918	lumbrius t

## ALIGNMENTS

RESULT 1	TKSL_AEDAE	STANDARD;	PRT;	10 AA.
ID	TKSL_AEDAE			
AC	P42634;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	STALOKININ I			
OS	Aedes aegypti (Yellowfever mosquito).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;			
OC	Aedes.			
OX	NCBI_Taxid=7159;			
RN	[1]			
RP	SEQUENCE			
RC	STRATIN-ROCKEFELLER, TISSUE-Salivary gland;			
RX	MEDLINE=94105119; PubMed=8278354;			
RA	Chapman D.E., Ribeiro J.M.C.;			
RT	"Stalokinin I and II: vasodilatory tachykinins from the yellow fever			
RT	mosquito Aedes aegypti."			
RL	Proc. Natl. Acad. Sci. U.S.A. 91:138-142(1994).			
CC	-I- FUNCTION: VASODILATORY PEPTIDE. MAY ACTIVATE MACROPHAGES AT THE			
CC	SITE OF FEEDING.			
CC	-I- SIMILARITY: SOME SIMILARITY TO TACHYKININS.			
CC	PIR; A49581; A49581.			
DR	InterPro; IPR002040; Tachykinin.			
DR	PROSITE; PS00267; TACHYKININ; 1.			
KW	Tachykinin; Neuropeptide; Amidation.			
FT	MOD_RES			
FT	SEQUENCE			
SQ	10 AA; 1145 MW; 3DCFDE6B59C33AA8 CRC64;			

Query Match 7.7%; Score 4; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	33 NTGD 36
Db	1 NTGD 4
RESULT 2	TKN2_SCYCA
ID	TKN2_SCYCA
AC	P08609;
DT	01-AUG-1988 (Rel. 08, Created)
DT	01-AUG-1988 (Rel. 08, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	SCYLIORHININ II (RECTIN).
OS	Scyllorhinus canicula (Spotted dogfish) (Spotted catshark).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC	Elasmobranchii; Galeomorphii; Galeoidea; Carcharhiniformes;
OC	Scyllorhinidae; Scyllorhinus.
OX	NCBI_Taxid=830;
RN	[1]
RP	SEQUENCE.

RX MEDLINE=86192829; PubMed=2422058;  
 RA Conlon J.M., Deacon C.F., O'Toole L., Thim L.;  
 RT "Scylorhinin I and II: two novel tachykinins from dogfish gut."  
 RL FEBS Lett. 200;111-116(1986).  
 RN [12]  
 RP SOURCE.  
 RC TISSUE=Small intestine;  
 RX MEDLINE=95335921; PubMed=7541963;  
 RA Anderson W.G., Conlon J.M., Hazon N.;  
 RT "Characterization of the endogenous intestinal peptide that  
 stimulates the rectal gland of scyliorhinus canicula."  
 RL Am. J. Physiol. 268:R1359-R1364(1995).  
 CC -I- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,  
 EVOKE BEHAVIORAL RESPONSES, ARE POTENT VASODILATORS AND  
 SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH  
 CC MUSCLES.  
 CC -I- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.  
 DR PIR: B24867; B24867.  
 DR Interpro: IPR002040; Tachykinin.  
 DR Pfam: PF02202; Tachykinin.1.  
 DR PROSITE: PS00267; TACHYKININ.1.  
 KW Tachykinin; Neuropeptide; Amidation.  
 FT DISULFID 7 13  
 FT MOD RES 18 18  
 SQ SEQUENCE 18 AA; 1854 MW; FCCA3PRLP2F3907 CRC64;

Query Match 7.7%; Score 4; DB 1; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PDGP 51  
 DB 8 PDGP 11

RESULT 3  
 ID ITHA\_PERAM STANDARD; PRT; 19 AA.  
 AC P19986;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE TREHALASE INHIBITOR (FRAGMENT).  
 OS Periplaneta americana (American cockroach).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
 OC Blattellidae; Blattidae; Periplaneta.  
 NCBI\_Taxid=6978;  
 RN [1]  
 RP SOURCE.  
 RC TISSUE=Hemolymph;  
 RX MEDLINE=89380218; PubMed=2777784;  
 RA Hayakawa Y., Jahagirdar A.P., Yaguchi M., Downer R.G.H.;  
 RT "Purification and characterization of trehalase inhibitor from  
 hemolymph of the American cockroach, Periplaneta americana."  
 RL J. Biol. Chem. 264:16165-16169(1989).  
 CC -I- FUNCTION: UNDER RESTING CONDITIONS, INHIBITS TREHALASE IN A  
 DOSE-DEPENDENT MANNER.  
 CC PIR: A34233; A34233.  
 DR GLYCOPROTEIN.  
 KW NON TER  
 FT NON TER  
 SQ SEQUENCE 19 AA; 2016 MW; A60C3A3397AF9A22 CRC64;

Query Match 7.7%; Score 4; DB 1; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPDG 50  
 DB 12 VPDG 15

RESULT 4  
 ID MIF\_PIG STANDARD; PRT; 20 AA.  
 AC P80928;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE MACROPHAGE MIGRATION INHIBITORY FACTOR (MIF) (PHENYLALANINE  
 TAUTOMERASE) (GLYCOSYLATION-INHIBITING FACTOR) (GIF) (FRAGMENT).  
 GN MIF.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 NCBI\_Taxid=9823;  
 RN [1]  
 RP SOURCE.  
 RA Riviere S., Bouet F., Menez A., Galat A.;  
 RL Submitted (MAR-1997) to the SWISS-PROT data bank.  
 CC -I- FUNCTION: THE EXPRESSION OF MIF AT SITES OF INFLAMMATION SUGGEST A  
 ROLE FOR THE MEDIATOR IN REGULATING THE FUNCTION OF MACROPHAGE IN  
 HOST DEFENSE. ALSO ACTS AS A PHENYLALANINE TAUTOMERASE (BY  
 CC SIMILARITY).  
 CC -I- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO THE MIF FAMILY.  
 DR HSP: P14174; MIF.  
 DR Interpro: IPR01398; MIF.  
 DR PROSITE: PS01158; MIF; PARTIAL.  
 KW Isomerase; Macrophage; Inflammatory response; Cytokine.  
 FT ACT SITE 1 1  
 FT NON TER 20 20  
 FT CATALYTIC BASE (BY SIMILARITY).  
 SQ SEQUENCE 20 AA; 2147 MW; 3517AF60F3012A61 CRC64;

Query Match 7.7%; Score 4; DB 1; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPDG 50  
 DB 14 VPDG 17

RESULT 5  
 ID EPHA8\_HUMAN STANDARD; PRT; 21 AA.  
 AC P29322;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE EPHRIN TYPE-A RECEPTOR 8 (EC 2.7.1.112) (TYROSINE-PROTEIN KINASE  
 DE RECEPTOR EEF) (EPH-AND ELK-RELATED KINASE) (HEK3) (FRAGMENT).  
 GN EPHA8 OR EEF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_Taxid=9606;  
 RN [1]  
 RP SOURCE FROM N.A.  
 RX MEDLINE=91296384; PubMed=1648701;  
 RA Chan J., Watt V.M.;  
 RT "eek and erk, new members of the eph subclass of receptor protein-  
 RT tyrosine kinases."  
 RL Oncogene 6:1057-1061(1991).  
 CC -I- FUNCTION: RECEPTOR FOR MEMBERS OF THE EPHRIN-A FAMILY.  
 CC -I- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -I- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. EPHRIN  
 CC RECEPTOR SUBFAMILY.  
 CC  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL: X59291; CAA41980.1; ALT\_INIT.  
DR PIR: S23361; S23361.  
DR MIM: 176945; -  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR001426; Receptor\_tyr\_kin\_V.  
DR InterPro: IPR001245; Tyr\_kin.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_AAP; PARTIAL.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; PARTIAL.  
DR PROSITE: PS00101; PROTEIN\_KINASE\_DOM; PARTIAL.  
DR PROSITE: PS00790; RECEPTOR\_TYR\_KIN\_V.1; PARTIAL.  
DR PROSITE: PS00791; RECEPTOR\_TYR\_KIN\_V.2; PARTIAL.  
KW transferase; tyrosine-protein kinase; ATP-binding; phosphorylation;  
KW Receptor.  
FT NON\_TER 1 1  
FT NON\_TER 21 21  
SQ SEQUENCE 21 AA; 2417 MW; 71922FDD46CA236C CRC64;

Query Match 7.7%; Score 4; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 ENGS 45  
Db 12 ENGS 15

RESULT 6  
PA22\_BOTSC STANDARD; PRT; 25 AA.  
AC P80963;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE PHOSPHOLIPASE A2 HOMOLOG, MYOTOXIN II (FRAGMENT).  
OS Bothriechis schlegelii (Evelash viper).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
OC Viperidae; Crotalinae; Bothriechis.  
OX NCBI\_TaxID=44725;  
RN [1]  
RP SEQUENCE.  
RC TISSUE-Venom;  
RX MEDLINE=97209510; PubMed=9056257;  
RA Angulo Y., Chaves E., Alape A., Rucavado A., Gutierrez J.M.,  
RT Lomonte B.;  
RT "Isolation and characterization of a myotoxic phospholipase A2 from  
RT the venom of the arboreal snake Bothriechis (Bothrops) schlegelii  
RT from Costa Rica.";  
RL Arch. Biochem. Biophys. 339:260-266(1997).  
CC -!- FUNCTION: THIS MYOTOXIC PROTEIN LACKS ENZYMATIC ACTIVITY.  
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.  
CC HSSP; P24605; ICLP.  
DR InterPro: IPR001211; PLP\_A2.  
DR Pfam: PF000068; Phoslip; 1.  
DR PROSITE: PS00118; PA2\_HIS; PARTIAL.  
DR PROSITE: PS00119; PA2\_ASP; PARTIAL.  
KW Toxin; Venom; Multigene family.  
FT NON\_TER 25 25  
SQ SEQUENCE 25 AA; 2725 MW; C578AB09235EA02A CRC64;

Query Match 7.7%; Score 4; DB 1; Length 25;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 TSYI 19

Db 19 TSYI 22

RESULT 7  
UBI1\_BOVIN STANDARD; PRT; 25 AA.  
AC P23356;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L1 (EC 3.4.19.12) (UCH-  
DE L1) (UBIQUITIN THIOLESTERASE L1) (NEURON CYTOPLASMIC PROTEIN 9.5)  
DE (PGP 9.5) (FRAGMENT).  
GN UCHL1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OX Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE.  
RC TISSUE-Brain;  
RX MEDLINE=92008646; PubMed=1833240;  
RA Giambanco I., Bianchi R., Ceccarelli P., Pula G., Sorci G.,  
RA Antonoli S., Bocchini V., Donato R.,  
RT "Neuron-specific protein gene product 9.5 (PGP 9.5) is also  
RT expressed in glioma cell lines and its expression depends on cellular  
RT growth state.";  
RL FEBS Lett. 290:131-134(1991).  
CC -!- FUNCTION: UBIQUITIN-PROTEIN HYDROLASE IS INVOLVED BOTH IN THE  
CC PROCESSING OF UBIQUITIN PRECURSORS AND OF UBIQUITINATED PROTEINS.  
CC THIS ENZYME IS A THIOLESTERASE THAT RECOGNIZE AND HYDROLYZE  
CC A PEPTIDE BOND AT THE C-TERMINAL GLYCINE OF UBIQUITIN.  
CC -!- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =  
CC UBIQUITIN + A THIOLE.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -!- TISSUE SPECIFICITY: NEURONS AND CELLS OF THE DIFFUSE  
CC NEUROENDOCRINE SYSTEM AND THEIR TUMORS.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C12; ALSO KNOWN AS  
CC FAMILY 1 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.  
DR PIR: S17561; S17561.  
DR MEROPS: C12.001; -.  
DR InterPro: IPR001578; UCH.  
DR PROSITE: PS00140; UCH\_1; PARTIAL.  
KW Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family.  
FT NON\_TER 25 25  
SQ SEQUENCE 25 AA; 2812 MW; 26B85ADD0A754D55 CRC64;

Query Match 7.7%; Score 4; DB 1; Length 25;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 DPEM 39  
Db 9 DPEM 12

RESULT 8  
AL21\_HORSE STANDARD; PRT; 29 AA.  
ID AL21\_HORSE  
AC P81216;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE DANDER MAJOR ALLERGEN EOU 2.0101 (FRAGMENT).  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE.

RC TISSUE-Skin;  
 MEDLINE=98237590; PubMed=9578478;  
 RA Bulone V., Krostad-Johnsen T., Smeestad-Paulsen B.;  
 RT "Separation of horse dander allergen proteins by two-dimensional  
 electrophoresis -- molecular characterization and identification of  
 RT Equ c 2.0101 and Equ c 2.0102 as lipocalin proteins.";  
 RT Eur. J. Biochem. 253:202-211(1998).  
 CC -1- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).  
 CC -1- PTM: N-GLYCOSYLATED.  
 CC -1- MISCELLANEOUS: PREDOMINANT ALLERGEN OF HORSE DANDER.  
 CC -1- SIMILARITY: BELONGS TO THE LIPOCALIN FAMILY.  
 CC -1- CAUTION: C.2.0101 AND C.2.0102 MIGHT BE TWO VARIANTS OF THE  
 CC SAME PROTEIN.  
 DR Interpro: IPR000566; Lipocalin\_cytfarp.  
 DR PROSITE: PS00213; LIPOCALIN: PARTIAL.  
 KW Transport: Lipocalin; Glycoprotein; Allergen.  
 FT NON\_TER 29 29  
 SQ SEQUENCE 29 AA; 3195 MW; 4BBED09FF17A4EC6 CRC64;

Query Match 7.7%; Score 4; DB 1; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 WNTI 15  
 III  
 DB 17 WNTI 20

RESULT 9  
 FARL\_HELP1 STANDARD; PRT; 7 AA.  
 ID FARL\_HELP1  
 AC P41871;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE FMRFAMIDE-LIKE NEUROPEPTIDE GDPPLRF-AMIDE.  
 OS Helisoma trivolvis (Snail).  
 CC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
 CC Planorbidae; Helisoma.  
 OX NCBI\_TaxID=27815;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-Kidney;  
 RX MEDLINE=94286417; PubMed=7912428;  
 RA Madrid K.P., Price D.A., Greenberg M.J., Khan H.R., Saleuddin A.S.M.;  
 RT "FMRFamide-related peptides from the kidney of the snail, Helisoma  
 trivolvis.";  
 RT Peptides 15:31-36(1994).  
 CC -1- FUNCTION: APPEARS TO BE INVOLVED IN OSMOREGULATION BY AFFECTING  
 CC THE KIDNEY, MANTEL AND SKIN.  
 CC -1- TISSUE SPECIFICITY: KIDNEY, SKIN, MANTEL AND THE HEMOLYMPH.  
 CC -1- SIMILARITY: BELONGS TO THE FMRFAMIDE RELATED PEPTIDE)  
 CC FAMILY.  
 KW Neuropeptide; Amidation.  
 FT MOD\_RES 7 7  
 SQ SEQUENCE 7 AA; 851 MW; 69D40729D76AA810 CRC64;

Query Match 5.8%; Score 3; DB 1; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 GDP 37  
 III  
 DB 1 GDP 3

RESULT 10  
 UF06\_MOUSE STANDARD; PRT; 8 AA.  
 ID UF06\_MOUSE  
 AC P38644;  
 DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE UNKNOWN PROTEIN FROM 2D-PAGE OF FIBROBLASTS (F50) (FRAGMENT).  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE-Fibroblast;  
 RX MEDLINE=9500907; PubMed=7523108;  
 RA Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;  
 RT "Separation and sequencing of familial and novel murine proteins  
 RT using preparative two-dimensional gel electrophoresis.";  
 RL Electrophoresis 15:735-745(1994).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 5.2, ITS MW IS: 50 KDA.  
 FT NON\_TER 8 8  
 SQ SEQUENCE 8 AA; 817 MW; A35DD878676B05B1 CRC64;

Query Match 5.8%; Score 3; DB 1; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PGG 8  
 III  
 DB 4 PGG 6

RESULT 11  
 CONO\_CONGE STANDARD; PRT; 9 AA.  
 ID CONO\_CONGE  
 AC P05486;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE LYS-CONOPRESSIN G.  
 OS Conus geographus (Geography cone).  
 CC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;  
 CC Neogastropoda; Conoidea; Conidae; Conus.  
 OX NCBI\_TaxID=6491;  
 RN [1]  
 RP SEQUENCE.  
 RC MEDLINE=88058932; PubMed=3680228;  
 RX Cruz L.J., de Santos V., Zafaralla G.C., Ramallo C.A., Zeikus R.D.,  
 RA Gray W.R., Olivera B.M.;  
 RT "Invertebrate vasopressin/oxytocin homologs. Characterization of  
 RT peptides from Conus geographus and Conus straitus venoms.";  
 RT J. Biol. Chem. 262:15821-15824(1987).  
 RN [2]  
 RP REVIEW.  
 RC MEDLINE=89024586; PubMed=3052286;  
 RA Gray W.R., Olivera B.M., Cruz L.J.;  
 RT "Peptide toxins from venomous Conus snails.";  
 RT Annu. Rev. Biochem. 57:665-700(1988).  
 CC -1- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.  
 DR PIR: A28495; A28495.  
 DR Interpro: IPR000981; Neuropeptides\_horm.  
 DR Pfam: PF00220; hormone4; 1.  
 DR PROSITE: PS00264; NEUROHYPOPHYS\_HORM; 1.  
 KW Hormone; Amidation.  
 FT DISULFID 1 6  
 FT MOD\_RES 9 9  
 SQ SEQUENCE 9 AA; 1037 MW; D4FC276B84540059 CRC64;

Query Match 5.8%; Score 3; DB 1; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 NCP 6  
 III



Db 5 NCP 7

RESULT 12

CONO\_CONST STANDARD: PRT: 9 AA.

AC P05487;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE ARG-CONOPRESSIN S.

OC Conus striatus (Striated cone).

OC Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;

OC Neogastropoda; Conoidea; Conidae; Conus.

OX NCBI\_TaxID=6493;

RN [1]

RP SEQUENCE

RX MEDLINE=88058932; PubMed=3680228;

RA Cruz L.J., de Santos V., Zafaralla G.C., Ramilo C.A., Zeikus R.D.,

RT "Invertebrate vasopressin/oxytocin homologs. Characterization of

RT peptides from Conus geographus and Conus striatus venoms."

RL J. Biol. Chem. 262:15821-15824(1987).

RN [2]

RP REVIEW.

RX MEDLINE=89024586; PubMed=3052286;

RA Gray W.R., Olivera B.M., Cruz L.J.;

RT "Peptide toxins from venomous Conus snails."

RL Annu. Rev. Biochem. 57:665-700(1988).

CC -1- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.

DR PIR: B28495; B28495.

DR InterPro: IPR000981; Neurohypophys\_horm.

DR Pfam: PF00220; hormone4; 1.

DR PROSITE: PS00264; NEUROHYPOPHYS\_HORM; 1.

KW Hormone; Amidation.

FT DISULFID 1 6

FT MOD\_RES 9 9

SO SEQUENCE 9 AA; 1031 MW; 17EB176EB4540050 CRC64;

Query Match 5.8%; Score 3; DB 1; Length 9;

Best Local Similarity 100.0%; Pred. No. 1e+05;

Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 NCP 6

Db 5 NCP 7

RESULT 13

DNFL\_LOCM1 STANDARD: PRT: 9 AA.

AC P16339;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE LOCUPRESSIN (DIURETIC NEUROPEPTIDE FL/F2).

OS Locusta migratoria (Migratory locust).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Orthopteroidea; Caellifera;

OC Acridomorpha; Acridoidea; Acrididae; Locusta.

OX NCBI\_TaxID=7004;

RN [1]

RP SEQUENCE

RC TISSUE=Subesophageal ganglion, and Thoracic ganglion;

RX MEDLINE=88077077; PubMed=3689410;

RA Proulx J.P., Miller C.A., Li J.P., Carney R.L., Girardie A.,

RT "Identification of an arginine vasopressin-like diuretic hormone from

RT Locusta migratoria."

RL Biochem. Biophys. Res. Commun. 149:180-186(1987).

CC -1- FUNCTION: DIURETIC HORMONE.

-1- SUBUNIT: F2 IS AN ANTIPARALLEL DISULFIDE LINKED DIMER OF FL.

CC -1- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.

DR PIR: A29477; A29477.

DR InterPro: IPR000981; Neurohypophys\_horm.

DR Pfam: PF00220; hormone4; 1.

DR PROSITE: PS00264; NEUROHYPOPHYS\_HORM; 1.

KW Hormone; Neuropeptide; Amidation.

FT DISULFID 1 6

FT DISULFID 1 1

FT DISULFID 6 6

FT MOD\_RES 9 9

SO SEQUENCE 9 AA; 976 MW; 56EB176EB451A057 CRC64;

Query Match 5.8%; Score 3; DB 1; Length 9;

Best Local Similarity 100.0%; Pred. No. 1e+05;

Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 NCP 6

Db 5 NCP 7

RESULT 14

ISOT\_CYPCA STANDARD: PRT: 9 AA.

AC P42993;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE ISOTOCIN.

OS Cyprinus carpio (Common carp).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

OC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.

OX NCBI\_TaxID=7962;

RN [1]

RP SEQUENCE.

RC TISSUE=pituitary;

RA Acher R., Chauvet J., Chauvet M.-T., Crepy D.;

RT "Characterization of neurohypophyseal hormones from a fresh water bony

RT fish, the carp (Cyprinus carpio). Comparison with hormones from sea

RT water bony fishs."

RL Comp. Biochem. Physiol. 14:245-254(1965).

CC -1- FUNCTION: ANTIDIURETIC HORMONE.

CC -1- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.

DR PIR: A61364; A61364.

DR InterPro: IPR000981; Neurohypophys\_horm.

DR Pfam: PF00220; hormone4; 1.

DR PROSITE: PS00264; NEUROHYPOPHYS\_HORM; 1.

KW Hormone; Pituitary.

FT DISULFID 1 6

FT MOD\_RES 9 9

SO SEQUENCE 9 AA; 969 MW; 17F476EB45B04B CRC64;

Query Match 5.8%; Score 3; DB 1; Length 9;

Best Local Similarity 100.0%; Pred. No. 1e+05;

Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 NCP 6

Db 5 NCP 7

RESULT 15

OXVA\_SCYCA STANDARD: PRT: 9 AA.

AC P42996;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE ASVATOCIN.

OS Scyliorhinus canicula (Spotted dogfish). (Spotted catshark).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 CC Elasmobranchii; Galeomorphi; Galeoidea; Carcharhiniformes;  
 CC Scyliorhinidae; Scyliorhinus.  
 OX NCBI\_TaxID=7830;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Plutitary; PubMed=7972045;  
 RX MEDLINE=95062247;  
 RA Chauvet J., Rouille Y., Chauveau C., Chauvet M.-T., Acher R.;  
 RT "Special evolution of neurohypophysial hormones in cartilaginous  
 fishes: asvatocin and phasvatocin, two oxytocin-like peptides  
 isolated from the spotted dogfish (Scyliorhinus caniculus).";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:11266-11270(1994).  
 CC -!- FUNCTION: DISPLAYS OXYTOCIC ACTIVITY ON RAT UTERUS.  
 CC -!- SIMILARITY: BELONGS TO THE VASOPRESSIN/OXYTOCIN FAMILY.  
 DR InterPro: IPR000981; Neurohypophys\_horm.  
 DR Pfam: PF00220; hormone4; 1.  
 DR PROSITE: PS00264; NEUROHYPOPHYS\_HORM; 1.  
 KW Hormone; Amidation.  
 FT DISULFID 1 6  
 FT MOD\_RES 9 9 AMIDATION.  
 SO SEQUENCE 9 AA; 982 MW; 17EDD76EB444404B CRC64;

Query Match 5.8%; Score 3; DB 1; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 NCP 6  
 DB 5 NCP 7

Search completed: March 28, 2002, 09:25:37  
 Job time: 535 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:00:11 ; Search time 62.93 Seconds

(without alignments)  
133.422 Million cell updates/sec

Title: US-09-726-348-2

Perfect score: 1243  
Sequence: 1 MAPHGPSLTITVPMALL.....TLVSILMATQRRAKTS 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127.5	10.3	2907	1 FBN2_MOUSE	061555 mus musculus
2	117.5	9.5	2911	1 FBN2_HUMAN	P35566 homo sapien
3	108.5	8.7	2871	1 FBN1_HUMAN	P35535 homo sapien
4	105	8.4	2703	1 NOTC_DROME	P07207 drosophila
5	102.5	8.2	1488	1 SERR_DROME	P18168 drosophila
6	102	8.2	1964	1 MTCA_MOUSE	P31695 mus musculus
7	101.5	8.2	2871	1 FBN1_MOUSE	061554 mus musculus
8	99.5	8.0	2871	1 FBN1_BOVIN	P98133 bos taurus
9	99	8.0	769	1 AD11_HUMAN	073078 homo sapien
10	99	8.0	773	1 AD11_MOUSE	091174 mus musculus
11	98.5	7.9	2437	1 NOTC_BRARE	P46530 brachydanio
12	98	7.9	1429	1 L112_CAEEL	P14585 caenorhabdi
13	98	7.9	2531	1 MTCL_MOUSE	001705 mus musculus
14	97.5	7.8	755	1 COMP_RAT	P35444 rattus norv
15	97.5	7.8	1712	1 TCGR_RAT	000918 rattus norv
16	96.5	7.8	2201	1 TENA_HUMAN	P24821 homo sapien
17	96	7.7	886	1 EMRI_HUMAN	Q14246 homo sapien
18	95.5	7.7	593	1 FA12_BOVIN	P98140 bos taurus
19	92.5	7.4	2444	1 NEIL_RAT	062919 rattus norv
20	92	7.4	2531	1 MTCL_HUMAN	P46531 homo sapien
21	92	7.4	2531	1 MTCL_RAT	007008 rattus norv
22	91.5	7.4	1394	1 TCGR_HUMAN	P22064 homo sapien
23	91.5	7.4	2318	1 MTCL_MOUSE	061982 sus scrofa
24	91.5	7.4	2871	1 FBN1_PIG	Q91936 sus scrofa
25	91	7.3	1114	1 TCGR_CHICK	P36197 gallus gall
26	90	7.2	2139	1 DLK_HUMAN	P80370 homo sapien
27	90	7.2	2139	1 CRB_DROME	P10040 drosophila
28	89.5	7.2	712	1 FBIL_CAEEL	077469 caenorhabdi
29	89.5	7.2	1217	1 EGF_MOUSE	P01133 mus musculus
30	89	7.2	365	1 DLK_MOUSE	Q09163 mus musculus
31	89	7.2	675	1 PRTS_RAT	P53813 rattus norv
32	88.5	7.1	638	1 Y053_HUMAN	P42331 homo sapien
33	88.5	7.1	1480	1 SLIT_DROME	P24014 drosophila

34	88.5	7.1	4660	1 LRP2_RAT	P98158 rattus norv
35	88	7.1	452	1 AD11_XENLA	Q96273 xenopus lae
36	88	7.1	1054	1 FBN1_STRPU	P10079 strongyloce
37	88	7.1	1295	1 GLP1_CAEEL	P13508 caenorhabdi
38	87.5	7.0	3097	1 CADN_DROME	Q15943 drosophila
39	87	7.0	746	1 MEPA_HUMAN	Q16819 homo sapien
40	87	7.0	2524	1 NOTC_XENLA	P21783 xenopus lae
41	86	6.9	473	1 FP2_MYNGA	Q25464 mytilus gal
42	86	6.9	757	1 COMP_HUMAN	P49747 homo sapien
43	85	6.8	810	1 NEIL_HUMAN	Q92832 homo sapien
44	84.5	6.8	833	1 DL_DROME	P10041 drosophila
45	83.5	6.7	402	1 LAG2_CAEEL	P45442 caenorhabdi

## ALIGNMENTS

```

RESULT 1
FBN2_MOUSE STANDARD: PRT: 2907 AA.
ID FBN2_MOUSE
AC 061555: 063957.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE FIBRILLIN 2 PRECURSOR.
GN FBN2 OR FBN-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95263670; PubMed=7744963;
RA Zhang H., Hu W., Ramirez F.,
RT "Developmental expression of fibrillin genes suggests heterogeneity
RT of extracellular microfibrils."
RL J. Cell Biol. 129:1165-1176(1995).
RN [2]
RP SEQUENCE OF 210-317 FROM N.A.
RX MEDLINE=94140368; PubMed=8307578;
RA Li X., Pereira L., Zhang H., Sanguineti C., Ramirez F., Bonadio J.,
RT Francke U.;
RT "Fibrillin genes map to regions of conserved mouse/human synteny on
RT mouse chromosomes 2 and 18."
RL Genomics 18:667-672(1993).
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: L39790; AAA74908.1; -
DR EMBL: S69359; AAC60685.1; -
DR MG: MG1:95490; Fbn2.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001438; EGF-TI.
DR InterPro: IPR002212; TB.
DR InterPro: IPR000822; ZnF-C2H2.
DR Pfam: PF00008; EGF; 46.
DR Pfam: PF00683; TB; 9.
DR PRINTS: PR00010; EGFBLD.
DR SMART: SM00179; EGF_CA; 43.

```





```

FT REPEAT 1572 1648 TGRFP 4.
FT DOMAIN 1649 1690 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1691 1732 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1733 1806 TGRFP 5.
FT DOMAIN 1807 1848 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1891 1932 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1933 1971 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1972 2014 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 2015 2054 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2055 2096 EGF-LIKE 35, CALCIUM-BINDING.
FT REPEAT 2097 2169 TGRFP 6.
FT DOMAIN 2170 2211 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2212 2251 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2252 2292 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2293 2336 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2337 2378 EGF-LIKE 40, CALCIUM-BINDING.
FT REPEAT 2379 2447 TGRFP 7.
FT DOMAIN 2448 2489 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2490 2530 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2531 2569 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2570 2612 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2613 2652 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2653 2693 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2694 2733 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 115 124 BY SIMILARITY.
FT DISULFID 119 130 BY SIMILARITY.
FT DISULFID 132 141 BY SIMILARITY.
FT DISULFID 149 159 BY SIMILARITY.
FT DISULFID 153 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 180 190 BY SIMILARITY.
FT DISULFID 184 195 BY SIMILARITY.
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 209 219 BY SIMILARITY.
FT DISULFID 286 300 BY SIMILARITY.
FT DISULFID 302 315 BY SIMILARITY.
FT DISULFID 321 333 BY SIMILARITY.
FT DISULFID 328 342 BY SIMILARITY.
FT DISULFID 344 357 BY SIMILARITY.
FT DISULFID 497 509 BY SIMILARITY.
FT DISULFID 504 518 BY SIMILARITY.
FT DISULFID 520 532 BY SIMILARITY.
FT DISULFID 538 548 BY SIMILARITY.
FT DISULFID 543 557 BY SIMILARITY.
FT DISULFID 559 572 BY SIMILARITY.
FT DISULFID 578 590 BY SIMILARITY.
FT DISULFID 585 599 BY SIMILARITY.
FT DISULFID 601 614 BY SIMILARITY.
FT DISULFID 620 631 BY SIMILARITY.
FT DISULFID 626 640 BY SIMILARITY.
FT DISULFID 642 655 BY SIMILARITY.
FT DISULFID 661 672 BY SIMILARITY.
FT DISULFID 667 681 BY SIMILARITY.
FT DISULFID 683 696 BY SIMILARITY.
FT DISULFID 771 783 BY SIMILARITY.
FT DISULFID 778 792 BY SIMILARITY.
FT DISULFID 794 807 BY SIMILARITY.
FT DISULFID 813 825 BY SIMILARITY.
FT DISULFID 820 834 BY SIMILARITY.
FT DISULFID 836 849 BY SIMILARITY.
FT DISULFID 855 865 BY SIMILARITY.
FT DISULFID 860 874 BY SIMILARITY.
FT DISULFID 876 889 BY SIMILARITY.
FT DISULFID 958 970 BY SIMILARITY.
FT DISULFID 965 979 BY SIMILARITY.
FT DISULFID 981 994 BY SIMILARITY.
FT DISULFID 1076 1088 BY SIMILARITY.
FT DISULFID 1093 1097 BY SIMILARITY.
FT DISULFID 1112 1117 BY SIMILARITY.
FT DISULFID 1118 1130 BY SIMILARITY.
FT DISULFID 1125 1139 BY SIMILARITY.
FT DISULFID 1141 1155 BY SIMILARITY.

```

```

FT DISULFID 1161 1173 BY SIMILARITY.
FT DISULFID 1168 1182 BY SIMILARITY.
FT DISULFID 1184 1197 BY SIMILARITY.
FT DISULFID 1203 1215 BY SIMILARITY.
FT DISULFID 1210 1224 BY SIMILARITY.

Query Match
Best Local Similarity 9.5%; Score 117.5; DB 1; Length 2911;
Matches 51; Conservative 18; Mismatches 77; Indels 77; Gaps 14;

QY 1 MAPHGPSLTTLVPMALLLALGVERALALPEICT--QCPSSVONLSKVAFCYKTTREL 58
Db 2435 ICPHGGYTTD-----GNDIDECKVMPNLCTNGC---INTMGSFRCFCVGYTT 2481
QY 59 MHAHRCCLMOKGTYLGLDQNSLEDPGP-NHQATTT-----VIIDLANPLKG 107
Db 2482 DISGTS-----IDDECS-QSPKPCNYTKTTESSYCCSPRGVLOEDSKTK- 2530
QY 108 DL-----ANFPGFTVQLTLTLPOHVNCPGINAMWNTITSYDNOICOGOK 153
Db 2531 DLDECOQKQHCQFLCVNLIGFT-----CKCPGFTQHH--TACIDNNCGSGP 2578
QY 154 NLGNMTGDPMECPENSGVDPGGLQVCADGFH---GYKC 192
Db 2579 LLCGKG---ICQNT-----PGSFCSCQRFSGFSDATGTLNC 2611

RESULT 3
ID FBNI_HUMAN STANDARD; PRT; 2871 AA.
AC P35555;
DT 01-JUN-1994 (Rel. 29, Created)
DT 20-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FIBRILIN 1 PRECURSOR.
GN FBNI OR FBN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93372860; PubMed=8364578;
RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA Pangillan T., Bonadio J.;
RT "Genomic organization of the sequence coding for fibrillin, the
RT defective gene product in Marfan syndrome.";
RL Hum. Mol. Genet. 2:961-968(1993).
RN [2]
RP SEQUENCE OF 1-932 FROM N.A.
RC TISSUE=Placenta, and Fibroblast;
RX MEDLINE=94010947; PubMed=7691719;
RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;
RA "Fibrillin binds calcium and is coded by cDNAs that reveal a
RA multidomain structure and alternatively spliced exons at the 5'
RA end.";
RL Nature 352:334-337(1991).
RN [3]
RP SEQUENCE OF 899-2871 FROM N.A.
RC MEDLINE=91304568; PubMed=1852207;
RA Maslen C.L., Corson G.M., Maddox B.K., Gianville R.W., Sakai L.Y.;
RA "Partial sequence of a candidate gene for the Marfan syndrome.";
RL Nature 352:334-337(1991).
RN [4]
RP SEQUENCE OF 813-1313 FROM N.A.
RC MEDLINE=91304567; PubMed=1852206;
RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
RA Tsipouras P., Ramirez F., Hollister D.W.;
RA "Linkage of Marfan syndrome and a phenotypically related disorder to
RA two different fibrillin genes.";
RL Nature 352:330-334(1991).
RN [5]

```

RP CHARACTERIZATION.  
 RA MEDLINE=91317849; PubMed=1860873;  
 RT Sakai L.Y., Keene D.R., Gianville R.W., Bachinger H.P.;  
 RT "Purification and partial characterization of fibrillin, a cysteine-  
 RT rich structural component of connective tissue microfibrils.";  
 RL J. Biol. Chem. 266:14763-14770(1991).  
 RN [6]  
 RP STRUCTURE BY NMR OF 2054-2125.  
 RX MEDLINE=98031893; PubMed=9362480;  
 RA Yuan X., Downing A.K., Knott V., Handford P.A.;  
 RT "Solution structure of the transforming growth factor beta-binding  
 RT protein-like module, a domain associated with matrix fibrils.";  
 RL EMBO J. 16:6659-6666(1997).  
 RN [7]  
 RP STRUCTURE BY NMR OF 2124-2205.  
 RX MEDLINE=96144829; PubMed=8568869;  
 RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;  
 RT "Calcium binding properties of an epidermal growth factor-like domain  
 RT pair from human fibrillin-1.";  
 RL J. Mol. Biol. 255:22-27(1996).  
 RN [8]  
 RP STRUCTURE BY NMR OF 2124-2205.  
 RX MEDLINE=96222301; PubMed=8653794;  
 RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,  
 RA Handford P.A.;  
 RT "Solution structure of a pair of calcium-binding epidermal growth  
 RT factor-like domains: implications for the Marfan syndrome and other  
 RT genetic disorders.";  
 RL Cell 85:597-605(1996).  
 RN [9]  
 RP REVIEW ON MFS VARIANTS.  
 RX MEDLINE=96174615; PubMed=8594563;  
 RA Collod G., Beroud C., Soussi T., Junten C., Boileau C.;  
 RT "Software and database for the analysis of mutations in the human  
 RT FBN1 gene.";  
 RL Nucleic Acids Res. 24:137-141(1996).  
 RN [10]  
 RP REVIEW ON MFS VARIANTS.  
 RX MEDLINE=97169383; PubMed=9016526;  
 RA Collod-Beroud G., Beroud C., Ades L., Black C., Boxer M., Brock D.J.,  
 RA Godfrey M., Hayward C., Karttunen L., Milewicz D., Peltonen L.,  
 RA Richards R.I., Wang W., Junten C., Boileau C.;  
 RT "Marfan Database (second edition): software and database for the  
 RT analysis of mutations in the human FBN1 gene.";  
 RL Nucleic Acids Res. 25:147-150(1997).  
 RN [11]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=96062175; PubMed=9401003;  
 RA Hayward C., Brock D.J.H.;  
 RT "Fibrillin-1 mutations in Marfan syndrome and other type-1  
 RT fibrillogenopathies.";  
 RL Hum. Mutat. 10:415-423(1997).  
 RN [12]  
 RP VARIANT MFS PRO-1137.  
 RX MEDLINE=91304569; PubMed=1852208;  
 RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,  
 RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,  
 RA Curristin S.M., Stetten G., Meyers D.A., Francomano C.A.;  
 RT "Marfan syndrome caused by a recurrent de novo missense mutation in  
 RT the fibrillin gene.";  
 RL Nature 352:337-339(1991).  
 RN [13]  
 RP VARIANTS MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.  
 RX MEDLINE=93250834; PubMed=1301946;  
 RA Dietz H.C., Sariva J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;  
 RT "Clustering of fibrillin (FBN) missense mutations in Marfan syndrome  
 RT patients at cysteine residues in EGF-like domains.";  
 RL Hum. Mutat. 1:366-374(1992).  
 RN [14]  
 RP VARIANT MFS SER-2307.  
 RX MEDLINE=92235290; PubMed=1569206;  
 RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J. Jr.,  
 RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;

RT "Marfan phenotype variability in a family segregating a missense  
 RT mutation in the epidermal growth factor-like motif of the fibrillin  
 RT gene.";  
 RL J. Clin. Invest. 89:1674-1680(1992).  
 RN [15]  
 RP VARIANTS MFS ILE-548 AND ALA-723.  
 RX MEDLINE=94010946; PubMed=8406497;  
 RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,  
 RA Pyeritz R.E., Francomano C.A.;  
 RT "Four novel FBN1 mutations: significance for mutant transcript level  
 RT and EGF-like domain calcium binding in the pathogenesis of Marfan  
 RT syndrome.";  
 RL Genomics 17:468-475(1993).  
 RN [16]  
 RP VARIANTS MFS SER-2144.  
 RX MEDLINE=93278402; PubMed=8504310;  
 RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;  
 RT "A novel fibrillin mutation in the Marfan syndrome which could  
 RT disrupt calcium binding of the epidermal growth factor-like module.";  
 RL Hum. Mol. Genet. 2:475-477(1993).  
 RN [17]  
 RP VARIANTS MFS R-862; Y-1117; P-1137 AND F-1589, AND VARIANT A-1148.  
 RX MEDLINE=94108431; PubMed=8281141;  
 RA Tynan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,  
 RA Berg M.A., Miller D.C., Francke U.;  
 RT "Mutation screening of complete fibrillin-1 coding sequence: report  
 RT of five new mutations, including two in 8-cysteine domains.";  
 RL Hum. Mol. Genet. 2:1813-1821(1993).  
 RN [18]  
 RP VARIANTS MFS GLY-217 AND ARG-2627.  
 RX MEDLINE=95067970; PubMed=7977366;  
 RA Karttunen L., Raghunath M., Loengqvist L., Peltonen L.;  
 RT "A compound heterozygous Marfan patient: two defective fibrillin  
 RT alleles result in a lethal phenotype.";  
 RL Am. J. Hum. Genet. 55:1083-1091(1994).  
 RN [19]  
 RP VARIANT EL LYS-2447.  
 RX MEDLINE=94245249; PubMed=8188302;  
 RA Longqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,  
 RA Peltonen L.;  
 RT "A novel mutation of the fibrillin gene causing ectopia lentis.";  
 RL Genomics 19:573-576(1994).  
 RN [20]  
 RP VARIANT MFS CYS-627.  
 RX MEDLINE=94272487; PubMed=8004112;  
 RA Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;  
 RT "Two novel mutations and a neutral polymorphism in EGF-like domains  
 RT of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan  
 RT syndrome patients.";  
 RL Hum. Mol. Genet. 3:373-375(1994).  
 RN [21]  
 RP VARIANT MFS CYS-122.  
 RX MEDLINE=94314977; PubMed=8040326;  
 RA Stahl-Hallengren C., Ukonen T., Kainulainen K., Kristoffersson U.,  
 RA Saxne T., Tornqvist K., Peltonen L.;  
 RT "An extra cysteine in one of the non-calcium-binding epidermal growth  
 RT factor-like motifs of the FBN1 polypeptide is connected to a novel  
 RT variant of Marfan syndrome.";  
 RL J. Clin. Invest. 94:709-713(1994).  
 RN [22]  
 RP VARIANT MFS TYR-1223.  
 RX MEDLINE=94351682; PubMed=8071963;  
 RA Hewett D.R., Lynch J.R., Child A., Sykes B.C.;  
 RT "A new missense mutation of fibrillin in a patient with Marfan  
 RT syndrome.";  
 RL J. Med. Genet. 31:338-339(1994).  
 RN [23]  
 RP VARIANT MFS HIS-1170.  
 RX MEDLINE=95174777; PubMed=7870075;  
 RA Hayward C., Porteous M.E.M., Brock D.J.H.;  
 RT "A novel mutation in the fibrillin gene (FBN1) in familial  
 RT arachnoidoclysis.";  
 RL Mol. Cell. Probes 8:325-327(1994).

```

RN [24]
RP VARIANTS MES G217;N1023;A1074;Y1242;R1513;E2127;W2151;K2447 AND R2511.
RX MEDLINE-94194368; PubMed-8136837;
RA Kainulainen K., Karttunen L., Puhakka L., Sakai L., Peltonen L.;
RT "Mutations in the fibrillin gene responsible for dominant ectopia
RT lentis and neonatal Marfan syndrome.";
RL Nat. Genet. 6:64-69(1994).
RN [25]
RP VARIANT SER-1127.
RX MEDLINE-95282774; PubMed-7762551;

Query Match      8.7%; Score 108.5; DB 1; Length 2871;
Best Local Similarity 22.1%; Pred. No. 0.096;
Matches 48; Conservative 18; Mismatches 72; Indels 79; Gaps 11;

QY 1 MAPHGPSLTLVPMWAAALLALGVERALALPEICT--QC---PGSYQNLKVA----- 49
DB 2389 LCPHGRGFMTN-----GADIDCKVYHDVCRNGECVNDGSHYCIKCTGYPPDIT 2438
QY 50 -----FYCKTRRLMLARCLNKGITLGLDLCNC-SLEDPGNFM 90
DB 2439 GTSCVDLNECNQAPKPCNFICKRNE---GSYQSCPKRGITLQEDGSKDLBCATKQH 2494
QY 91 QAHTTVIIDLQANPLKGDLANTFPRGTLOTLILPOHVNGPFGGINAMNTTSTYIDNOICQ 150
DB 2495 NCQFLCV-----NTIGGF-----CKCPPTQTHH--TSCIDNNECT 2529
QY 151 GOKNLNNTGDPENCPCNGSCVPDGPGLQCVCADGF 187
DB 2530 SDINLCGSKG---ICQNT-----PQSFTCECORGF 2556

RESULT 4
NOTC_DROME STANDARD; PRT; 2703 AA.
ID NOTC_DROME STANDARD; PRT; 2703 AA.
AC P07207; P04154;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-86079539; PubMed-3935325;
RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT "Nucleotide sequence from the neurogenic locus notch implies a gene
RT product that shares homology with proteins containing EGF-like
RT repeats.";
RL Cell 43:567-581(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-OREGON-R;
RX MEDLINE-87064624; PubMed-3097517;
RA Kidd S., Kelley M.R., Young M.W.;
RT "Sequence of the notch locus of Drosophila melanogaster: relationship
RT of the encoded protein to mammalian clotting and growth factors.";
RL Mol. Cell. Biol. 6:3094-3108(1986).
RN [3]
RP SEQUENCE OF 2505-2611 FROM N.A.
RX MEDLINE-85093929; PubMed-2981631;
RA Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.;
RT "opa: a novel family of transcribed repeats shared by the Notch locus
RT and other developmentally regulated loci in D. melanogaster.";
RL Cell 40:55-62(1985).
RN [4]
RP SEQUENCE OF 1-8 FROM N.A.
RX MEDLINE-87257846; PubMed-3037327;
RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;
RT "Restriction of P-element insertions at the Notch locus of Drosophila
RT melanogaster.";
RL Mol. Cell. Biol. 7:1545-1548(1987).
RN [5]
RP REVIEW.
RA Harris W.A.;
RT "Many cell types specified by Notch function.";
RL Curr. Biol. 1:120-122(1991).
CC -1- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
CC ECTODERM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
CC NEUROGENIC GENES.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16152; AAB59220.1; -.
DR EMBL; M16153; AAB59220.1; JOINED.
DR EMBL; M16149; AAB59220.1; JOINED.
DR EMBL; M16150; AAB59220.1; JOINED.
DR EMBL; M16151; AAB59220.1; JOINED.
DR EMBL; K03508; AAA28725.1; -.
DR EMBL; M13689; AAA28725.1; JOINED.
DR EMBL; K03507; AAA28725.1; JOINED.
DR EMBL; M12175; AAA74496.1; -.
DR EMBL; M16025; AAA28726.1; -.
DR PIR; A24420; A24420.
DR PIR; A24768; A24768.
DR PIR; A05267; A05267.
DR HSSP; P00740; 11XA.
DR Flybase; FBgn004647; N.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFLOOD.
DR SMART; SM00248; ANK; 4.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00001; EGF-like; 13.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 28.
DR PROSITE; PS01187; EGF_CA; 22.
DR Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
KM SIGNL. 1 44
FT CHAIN 45 2703
FT DOMAIN 45 1745 NEUROGENIC LOCUS NOTCH PROTEIN.
FT TRANSMEM 1746 1766 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1767 2703 POTENTIAL.
FT DOMAIN 1767 2703 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 58 95 EGF-LIKE 1.
FT DOMAIN 96 136 EGF-LIKE 2.

```



```

FT DOMAIN 139 176 EGF-LIKE 3.
FT DOMAIN 177 215 EGF-LIKE 4.
FT DOMAIN 217 253 EGF-LIKE 5.
FT DOMAIN 255 291 EGF-LIKE 6.
FT DOMAIN 293 329 EGF-LIKE 7.
FT DOMAIN 331 370 EGF-LIKE 8.
FT DOMAIN 372 408 EGF-LIKE 9.
FT DOMAIN 409 447 EGF-LIKE 10.
FT DOMAIN 448 486 EGF-LIKE 11.
FT DOMAIN 488 524 EGF-LIKE 12.
FT DOMAIN 526 562 EGF-LIKE 13.
FT DOMAIN 564 600 EGF-LIKE 14.
FT DOMAIN 602 637 EGF-LIKE 15.
FT DOMAIN 639 673 EGF-LIKE 16.
FT DOMAIN 677 713 EGF-LIKE 17.
FT DOMAIN 715 751 EGF-LIKE 18.
FT DOMAIN 753 789 EGF-LIKE 19.
FT DOMAIN 791 827 EGF-LIKE 20.
FT DOMAIN 829 865 EGF-LIKE 21.
FT DOMAIN 867 905 EGF-LIKE 22.
FT DOMAIN 907 944 EGF-LIKE 23.
FT DOMAIN 946 982 EGF-LIKE 24.
FT DOMAIN 1020 1058 EGF-LIKE 25.
FT DOMAIN 1062 1096 EGF-LIKE 26.
FT DOMAIN 1098 1134 EGF-LIKE 27.
FT DOMAIN 1136 1181 EGF-LIKE 28.
FT DOMAIN 1183 1219 EGF-LIKE 29.
FT DOMAIN 1221 1257 EGF-LIKE 30.
FT DOMAIN 1259 1295 EGF-LIKE 31.
FT DOMAIN 1297 1335 EGF-LIKE 32.
FT DOMAIN 1337 1373 EGF-LIKE 33.
FT DOMAIN 1375 1412 EGF-LIKE 34.
FT DOMAIN 1415 1451 EGF-LIKE 35.
FT DOMAIN 1475 1513 EGF-LIKE 36.
FT REPEAT 1514 1553 LIN/NOTCH 1.
FT REPEAT 1554 1593 LIN/NOTCH 2.
FT REPEAT 1594 1633 LIN/NOTCH 3.
FT REPEAT 1634 1673 ANK 1.
FT REPEAT 1674 1713 ANK 2.
FT REPEAT 1714 1753 ANK 3.
FT REPEAT 1754 1793 ANK 4.
FT REPEAT 1794 1833 ANK 5.
FT REPEAT 1834 1873 ANK 6.
FT REPEAT 1874 1913 POLY-GLN (OPA-REPEAT).
FT REPEAT 1914 1953 BY SIMILARITY.
FT REPEAT 1954 1993 BY SIMILARITY.
FT REPEAT 1994 2033 BY SIMILARITY.
FT REPEAT 2034 2073 BY SIMILARITY.
FT REPEAT 2074 2113 BY SIMILARITY.
FT REPEAT 2114 2153 BY SIMILARITY.
FT REPEAT 2154 2193 BY SIMILARITY.
FT REPEAT 2194 2233 BY SIMILARITY.
FT REPEAT 2234 2273 BY SIMILARITY.
FT REPEAT 2274 2313 BY SIMILARITY.
FT REPEAT 2314 2353 BY SIMILARITY.
FT REPEAT 2354 2393 BY SIMILARITY.
FT REPEAT 2394 2433 BY SIMILARITY.
FT REPEAT 2434 2473 BY SIMILARITY.
FT REPEAT 2474 2513 BY SIMILARITY.
FT REPEAT 2514 2553 BY SIMILARITY.
FT REPEAT 2554 2593 BY SIMILARITY.
FT REPEAT 2594 2633 BY SIMILARITY.
FT REPEAT 2634 2673 BY SIMILARITY.
FT REPEAT 2674 2713 BY SIMILARITY.
FT REPEAT 2714 2753 BY SIMILARITY.
FT REPEAT 2754 2793 BY SIMILARITY.
FT REPEAT 2794 2833 BY SIMILARITY.
FT REPEAT 2834 2873 BY SIMILARITY.
FT REPEAT 2874 2913 BY SIMILARITY.
FT REPEAT 2914 2953 BY SIMILARITY.
FT REPEAT 2954 2993 BY SIMILARITY.
FT REPEAT 2994 3033 BY SIMILARITY.
FT REPEAT 3034 3073 BY SIMILARITY.
FT REPEAT 3074 3113 BY SIMILARITY.
FT REPEAT 3114 3153 BY SIMILARITY.
FT REPEAT 3154 3193 BY SIMILARITY.
FT REPEAT 3194 3233 BY SIMILARITY.
FT REPEAT 3234 3273 BY SIMILARITY.
FT REPEAT 3274 3313 BY SIMILARITY.
FT REPEAT 3314 3353 BY SIMILARITY.
FT REPEAT 3354 3393 BY SIMILARITY.
FT REPEAT 3394 3433 BY SIMILARITY.
FT REPEAT 3434 3473 BY SIMILARITY.
FT REPEAT 3474 3513 BY SIMILARITY.
FT REPEAT 3514 3553 BY SIMILARITY.
FT REPEAT 3554 3593 BY SIMILARITY.
FT REPEAT 3594 3633 BY SIMILARITY.
FT REPEAT 3634 3673 BY SIMILARITY.
FT REPEAT 3674 3713 BY SIMILARITY.
FT REPEAT 3714 3753 BY SIMILARITY.
FT REPEAT 3754 3793 BY SIMILARITY.
FT REPEAT 3794 3833 BY SIMILARITY.
FT REPEAT 3834 3873 BY SIMILARITY.
FT REPEAT 3874 3913 BY SIMILARITY.
FT REPEAT 3914 3953 BY SIMILARITY.
FT REPEAT 3954 3993 BY SIMILARITY.
FT REPEAT 3994 4033 BY SIMILARITY.
FT REPEAT 4034 4073 BY SIMILARITY.
FT REPEAT 4074 4113 BY SIMILARITY.
FT REPEAT 4114 4153 BY SIMILARITY.
FT REPEAT 4154 4193 BY SIMILARITY.
FT REPEAT 4194 4233 BY SIMILARITY.
FT REPEAT 4234 4273 BY SIMILARITY.
FT REPEAT 4274 4313 BY SIMILARITY.
FT REPEAT 4314 4353 BY SIMILARITY.

```

```

FT DISULFID 437 446 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 459 474 BY SIMILARITY.
FT DISULFID 476 485 BY SIMILARITY.
FT DISULFID 492 503 BY SIMILARITY.
FT DISULFID 497 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 568 579 BY SIMILARITY.
FT DISULFID 573 588 BY SIMILARITY.
FT DISULFID 590 599 BY SIMILARITY.
FT DISULFID 606 616 BY SIMILARITY.
FT DISULFID 611 625 BY SIMILARITY.

Query Match
Best local Similarity 24.6%; Score 105; DB 1; Length 2703;
Matches 47; Conservative 17; Mismatches 59; Indels 68; Gaps 10;

QY 27 RALALPEICTGCGSVQNLKFA-----YCK-----TRELMLAR 63
Db 1147 KGLSLKQLCNMGTCRDYNSHWCVCSQGYAGSCQKEIDECOSQPCQNGCTRDIGAVE 1206
QY 64 CCLNOKGTTILGLDLCNSTEDPGNPHQAHTTYIIDLANPLK--GDLANTFRGFTQLOT 121
Db 1207 CQCRQ-----GRGQNCLEL-----IDDCAPNFCQNGCTCHDRVMNS----- 1244
QY 122 LILPQHVNCPGGINAWNTTYSYIDNOCIGORNLNNTGDPMPCEPENGSCVPDGLQOC 181
Db 1245 -----CSCPFG-----TWGI-----ICEINKDDCK-----PCACHNNGSCI-DVYGGEFC 1283
QY 182 VCADGPHGYKC 192
Db 1284 VCPGFEVGARC 1294

RESULT 5
SERR_DROME STANDARD; PRT; 1408 AA.
AC P18168;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE SERRATE PROTEIN PRECURSOR (BRADED PROTEIN).
GN SER OR BD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=91347903; PubMed=1840519;
RA Thomas U., Speicher S.A., Knust E.;
RT "The Drosophila gene Serrate encodes an EGF-like transmembrane
protein with a complex expression pattern in embryos and wing
discs.";
RT discs";
RL Development 111:749-761(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91099666; PubMed=2125287;
RA Fleming R.J., Scottgale T.N., Diederich R.J., Artavanis-Tsakonas S.;
RT "The gene Serrate encodes a putative EGF-like transmembrane protein
essential for proper ectodermal development in Drosophila
melanogaster.";
RT Genes Dev. 4:2188-2201(1990).
CC -!- FUNCTION: ESSENTIAL FOR PROPER ECTODERMAL DEVELOPMENT. SERRATE
CC MAY REPRESENT AN ELEMENT IN A NETWORK OF INTERACTING MOLECULES
CC OPERATING AT THE CELL SURFACE DURING THE DIFFERENTIATION OF
CC CERTAIN TISSUES.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

```





```

FT DISULFID 433 449 BY SIMILARITY.
FT DISULFID 443 458 BY SIMILARITY.
FT DISULFID 460 469 BY SIMILARITY.
FT DISULFID 476 487 BY SIMILARITY.
FT DISULFID 481 496 BY SIMILARITY.
FT DISULFID 498 507 BY SIMILARITY.
FT DISULFID 514 525 BY SIMILARITY.
FT DISULFID 519 534 BY SIMILARITY.
FT DISULFID 536 545 BY SIMILARITY.
FT DISULFID 552 563 BY SIMILARITY.
FT DISULFID 557 572 BY SIMILARITY.
FT DISULFID 574 583 BY SIMILARITY.
FT DISULFID 590 601 BY SIMILARITY.
FT DISULFID 595 610 BY SIMILARITY.
FT DISULFID 612 621 BY SIMILARITY.
FT DISULFID 626 637 BY SIMILARITY.
FT DISULFID 631 646 BY SIMILARITY.
FT DISULFID 648 655 BY SIMILARITY.
FT DISULFID 662 674 BY SIMILARITY.
FT DISULFID 676 685 BY SIMILARITY.
FT DISULFID 692 703 BY SIMILARITY.
FT DISULFID 697 712 BY SIMILARITY.
FT DISULFID 714 723 BY SIMILARITY.
FT DISULFID 730 741 BY SIMILARITY.
FT DISULFID 735 750 BY SIMILARITY.
FT DISULFID 752 761 BY SIMILARITY.
FT DISULFID 768 779 BY SIMILARITY.
FT DISULFID 773 788 BY SIMILARITY.
FT DISULFID 790 799 BY SIMILARITY.
FT DISULFID 807 818 BY SIMILARITY.
FT DISULFID 812 827 BY SIMILARITY.
FT DISULFID 829 838 BY SIMILARITY.
FT DISULFID 845 856 BY SIMILARITY.
FT DISULFID 850 865 BY SIMILARITY.
FT DISULFID 867 876 BY SIMILARITY.
FT DISULFID 882 893 BY SIMILARITY.
FT DISULFID 897 912 BY SIMILARITY.
FT DISULFID 914 923 BY SIMILARITY.
FT DISULFID 930 941 BY SIMILARITY.
FT DISULFID 935 950 BY SIMILARITY.
FT DISULFID 952 961 BY SIMILARITY.
FT DISULFID 966 979 BY SIMILARITY.
FT DISULFID 973 988 BY SIMILARITY.
FT DISULFID 990 999 BY SIMILARITY.
FT DISULFID 1006 1019 BY SIMILARITY.
FT DISULFID 1011 1028 BY SIMILARITY.
FT DISULFID 1030 1039 BY SIMILARITY.
FT DISULFID 1046 1057 BY SIMILARITY.
FT DISULFID 1051 1066 BY SIMILARITY.
FT DISULFID 1071 1080 BY SIMILARITY.
FT DISULFID 1087 1098 BY SIMILARITY.

```

Query Match 8.2%; Score 102; DB 1; Length 1964;

Best Local Similarity 21.1%; Pred. No. 0.25; Mismatches 75; Indels 80; Gaps 11;

```

Matches 48; Conservative 24;
QY 35 CTGCGPVQNLKVAFCCKTRELMLHARC-----CLNKGTLIGL-----DLQ 78
DB 750 CTCLP-----SHTRGHCQAVDHCVASCLNGCTCVNKRGTFFCLCATGFGGLHCEKT 803
QY 79 NCSLEDP-----GPNFQAHTTVIID-----LQANPLK 106
DB 904 NPSCADSPCKNKATCQDTPRGARCLSPGTYGSSCQTLIDLCARKPCPHARCQSGSPF 863
QY 107 GDLANTEFGFT-----OLQTLILPOHVN-----CGGIANWNTTTSY-----IDNQ 147
DB 864 QCLC--LQGTGALCDPLPLSCQMAAMSGRIETSLCONGSLCIDTGSSTFCRCPFPQGR 921
QY 148 ICQGNKILCNNTGDPKCPENGSCVPPDGPGLLOCVACADGFHYGYCMR 194
DB 922 LCONMNPNC-----BPNPCHGSTCVPOPSGYV--CQACAGYEGONGSK 963

```

```

RESULT 7
ID FBN1_MOUSE STANDARD: PRT: 2871 AA.
AC 061554; Q60826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE FIBRILIN 1 PRECURSOR.
GN FBN1 OR FBN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=95130561; PubMed=7829516;
RA Yin W., Germiller J., Sanguineti C., Smiley E., Panglilman T.,
RA Pereira L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene."
RL J. Biol. Chem. 270:1798-1806(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RA Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -1- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC MICROFIBRILS (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L29454; AAA56840.1;
DR EMBL: U22493; AAA64217.1;
DR HSSP: P35555; IAPJ.
DR MGD: MGI:95489; Fbn1.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002212; TB.
DR Pfam: PF00008; EGF; 46.
DR Pfam: PF00683; TB; 9.
DR SMART: SM00179; EGF_CA; 42.
DR SMART: SM00010; EGF_Like; 4.
DR PROSITE: PS00010; ASX_HYDROXYL; 43.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS0186; EGF_2; 38.
DR PROSITE: PS0187; EGF_CA; 45.
KW Extracellular matrix; calcium-binding; glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 1 27
FT CHAIN 28 2871 FIBRILIN 1.
FT DOMAIN 81 112 EGF-LIKE 1, NON-CALCIUM BINDING.
FT DOMAIN 115 146 EGF-LIKE 2, NON-CALCIUM BINDING.
FT DOMAIN 147 178 EGF-LIKE 3, NON-CALCIUM BINDING.
FT DOMAIN 246 287 EGF-LIKE 4, CALCIUM-BINDING.
FT DOMAIN 288 329 EGF-LIKE 5, CALCIUM-BINDING.
FT DOMAIN 330 401 TGFBP 1.
FT DOMAIN 402 446 PRO-RICH.
FT DOMAIN 449 489 EGF-LIKE 6, NON-CALCIUM BINDING.

```

```
FT DOMAIN 490 529 EGF-LIKE 7, CALCITRIUM-BINDING.
FT DOMAIN 530 571 EGF-LIKE 8, CALCITRIUM-BINDING.
FT DOMAIN 572 612 EGF-LIKE 9, CALCITRIUM-BINDING.
FT DOMAIN 613 653 EGF-LIKE 10, CALCITRIUM-BINDING.
FT DOMAIN 656 721 TGFBR 2.
FT DOMAIN 723 764 EGF-LIKE 11, CALCITRIUM-BINDING.
FT DOMAIN 765 806 EGF-LIKE 12, CALCITRIUM-BINDING.
FT DOMAIN 807 846 EGF-LIKE 13, CALCITRIUM-BINDING.
FT DOMAIN 910 951 EGF-LIKE 14, CALCITRIUM-BINDING.
FT DOMAIN 952 1018 TGFBR 3.
FT DOMAIN 1028 1069 EGF-LIKE 15, CALCITRIUM-BINDING.
FT DOMAIN 1070 1112 EGF-LIKE 16, CALCITRIUM-BINDING.
FT DOMAIN 1113 1154 EGF-LIKE 17, CALCITRIUM-BINDING.
FT DOMAIN 1155 1196 EGF-LIKE 18, CALCITRIUM-BINDING.
FT DOMAIN 1197 1237 EGF-LIKE 19, CALCITRIUM-BINDING.
FT DOMAIN 1238 1279 EGF-LIKE 20, CALCITRIUM-BINDING.
FT DOMAIN 1280 1321 EGF-LIKE 21, CALCITRIUM-BINDING.
FT DOMAIN 1322 1362 EGF-LIKE 22, CALCITRIUM-BINDING.
FT DOMAIN 1363 1403 EGF-LIKE 23, CALCITRIUM-BINDING.
FT DOMAIN 1404 1445 EGF-LIKE 24, CALCITRIUM-BINDING.
FT DOMAIN 1446 1486 EGF-LIKE 25, CALCITRIUM-BINDING.
FT DOMAIN 1487 1527 EGF-LIKE 26, CALCITRIUM-BINDING.
FT DOMAIN 1528 1599 TGFBR 4.
FT DOMAIN 1606 1647 EGF-LIKE 27, CALCITRIUM-BINDING.
FT DOMAIN 1648 1688 EGF-LIKE 28, CALCITRIUM-BINDING.
FT DOMAIN 1689 1758 TGFBR 5.
FT DOMAIN 1766 1807 EGF-LIKE 29, CALCITRIUM-BINDING.
FT DOMAIN 1808 1848 EGF-LIKE 30, CALCITRIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 31, CALCITRIUM-BINDING.
FT DOMAIN 1891 1929 EGF-LIKE 32, CALCITRIUM-BINDING.
FT DOMAIN 1930 1972 EGF-LIKE 33, CALCITRIUM-BINDING.
FT DOMAIN 1973 2012 EGF-LIKE 34, CALCITRIUM-BINDING.
FT DOMAIN 2013 2054 EGF-LIKE 35, CALCITRIUM-BINDING.
FT DOMAIN 2055 2121 TGFBR 6.
FT DOMAIN 2127 2165 EGF-LIKE 36, CALCITRIUM-BINDING.
FT DOMAIN 2166 2205 EGF-LIKE 37, CALCITRIUM-BINDING.
FT DOMAIN 2206 2246 EGF-LIKE 38, CALCITRIUM-BINDING.
FT DOMAIN 2247 2290 EGF-LIKE 39, CALCITRIUM-BINDING.
FT DOMAIN 2291 2332 EGF-LIKE 40, CALCITRIUM-BINDING.
FT DOMAIN 2333 2400 TGFBR 7.
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCITRIUM-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 42, CALCITRIUM-BINDING.
FT DOMAIN 2485 2524 EGF-LIKE 43, CALCITRIUM-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCITRIUM-BINDING.
FT DOMAIN 2567 2606 EGF-LIKE 45, CALCITRIUM-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCITRIUM-BINDING.
FT DOMAIN 2648 2687 EGF-LIKE 47, CALCITRIUM-BINDING.
FT DOMAIN 85 94 BY SIMILARITY.
FT DISULFID 89 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
FT DISULFID 119 129 BY SIMILARITY.
FT DISULFID 123 134 BY SIMILARITY.
FT DISULFID 136 145 BY SIMILARITY.
FT DISULFID 150 160 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 250 262 BY SIMILARITY.
FT DISULFID 257 271 BY SIMILARITY.
FT DISULFID 273 286 BY SIMILARITY.
FT DISULFID 292 304 BY SIMILARITY.
FT DISULFID 299 313 BY SIMILARITY.
FT DISULFID 315 328 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 460 474 BY SIMILARITY.
FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
```

```
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1389 1402 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.
```

Query Match Best Local Similarity 8.2%; Score 101.5; DB 1; Length 2871;

Matches 47; Conservative 18; Mismatches 73; Indels 79; Gaps 11;

```
QY 1 MAPHGSGITTTVPWMAALLALGVERALALPEICT--QC---PGSVGNLSKVA----- 49
Db 2389 LCPHGGGFTN-----GADVDECKVIHDVCRNCGVNDKSTHCTGTGTPDIT 2438
QY 50 -----FYCKTRELMLHARCCLNKGTLLGLDLNC-SLEDPGPNFH 90
Db 2439 GTSVDLNECNDQAPKPCNFKNTE---GSYQSCSPGNYILQEDGRCKDLDECAIKQH 2494
QY 91 QAHTVIYIDLQANPLKGLDANFRGFTQLOTLPLPQAHNCGGGINANNVTTSYIDNICO 150
Db 2495 NCQPLCV-----MTIGFT-----CKPPGFTQHH--TACIDNNECT 2529
```



```

FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 776 790 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 830 845 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.

Query Match 8.08; Score 99.5; DB 1; Length 2871;
Best Local Similarity 21.2%; Pred. No. 0.63;
Matches 46; Conservative 19; Mismatches 73; Indels 79; Gaps 11;

```

```

Db 2439 GTACVDLMECNQAPRCPNCFKNTNE----GSYQSCPAFYLLQEDGRCCKDLDECAATKQH 2494
Qy 91 QAHYIVIDLQANPLKGLDANFREGFOTLOLLIPQHNCGGGINAMNTTISYDNOICQ 150
Db 2495 NQGFICV-----NTIGSF-----CKCPGFQHH--TACIDNNECT 2529
Qy 151 GOKNLCNNTGPEMCPENGSCVPDPGILQCVACDGF 187
Db 2530 SDINIGSKG---ICQNT-----PGSFTECCQRGF 2556

RESULT 9
AD11_HUMAN STANDARD; PRT; 769 AA.
AC 075078; Q14808; Q14809; Q14810;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADAM 11 PRECURSOR (A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 11)
DE (METALLOPROTEINASE-LIKE, DISINTEGRIN-LIKE, AND CYSTEINE-RICH PROTEIN)
DE (MDC).
GN ADAM11 OR MDC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=98359734; PubMed=9693107;
RA Sagane K., Ohya Y., Hasegawa Y., Tanaka I.;
RT "Metalloproteinase-like, disintegrin-like, cysteine-rich proteins MDC2
RT and MDC3: novel human cellular disintegrins highly expressed in the
RT brain.";
RL Biochem. J. 334:93-98(1998).
RN [2]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Cerebellum;
RX MEDLINE=94073190; PubMed=8252040;
RA Emi M., Katagiri T., Harada Y., Saito H., Inazawa J., Ito I.,
RA Kasumi F., Nakamura Y.;
RT "A novel metalloproteinase/disintegrin-like gene at 17q21.3 is
RT somatically rearranged in two primary breast cancers.";
RL Nat. Genet. 5:151-157(1993).
RN [3]
RP SEQUENCE OF 106-769 FROM N.A. (LONG AND SHORT ISOFORMS).
RC TISSUE=Brain, Ovary, Testis, and Breast;
RX MEDLINE=95044425; PubMed=7956356;
RA Katagiri T., Harada Y., Emi M., Nakamura Y.;
RT "Human metalloproteinase/disintegrin-like (MDC) gene: exon-intron
RT organization and alternative splicing.";
RL Cytogenet. Cell Genet. 68:39-44(1995).
CC -1- FUNCTION: PROBABLE LIGAND FOR INTEGRIN IN THE BRAIN. THIS IS A NON
CC CATALYTIC METALLOPROTEINASE-LIKE PROTEIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM/MDC-769 (SHOWN HERE)
CC AND A SHORT FORM/MDC-524. ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN BRAIN. SLIGHTLY
CC DETECTED OR NOT AT ALL IN OTHER TISSUES.
CC -1- DOMAIN: A CONSERVED MOTIF [AVN(E/D)CD] WITHIN THE DISINTEGRIN-LIKE
CC DOMAIN COULD BE INVOLVED IN THE BINDING TO THE INTEGRIN RECEPTOR.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUZIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- DISEASE: REARRANGEMENTS OCCUR IN BREAST AND OVARIAN CANCERS WHICH
CC INVOLVE MULTIPLE EXONS AND DISRUPT THE CODING REGION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC METALLOPROTEINASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AB009675; BAA32352.1; -  
 CC EMBL: D17390; BAA04213.1; -  
 CC EMBL: D31872; BAA06670.1; -  
 CC EMBL: D31872; BAA06671.1; -  
 CC HSSP: P18619; 1FVL.  
 CC MEROPS: M12.976; -  
 CC MIM: J15120; -  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001590; Reprolysin.  
 DR InterPro: IPR001762; Disintegrin.  
 DR InterPro: IPR002870; Pep\_M12B\_propep.  
 DR Pfam: PF00200; disintegrin\_1.  
 DR Pfam: PF01562; Rep\_M12B\_propep; 1.  
 DR Pfam: PF01421; Repolysin; 1.  
 DR ProDom: PD000664; disintegrin; 1.  
 DR SMART: SM00050; DISIN; 1.  
 DR SMART: SM00181; EGF; 1.  
 DR PROSITE: PS00215; ADAM\_MERPO; 1.  
 DR PROSITE: PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR PROSITE: PS00214; DISINTEGRIN\_2; 1.  
 DR PROSITE: PS00022; EGF\_1; 1.  
 DR PROSITE: PS01186; EGF\_2; FALSE\_NEG.  
 DR PROSITE: PS00142; ZINC\_PROTEASE; FALSE\_NEG.  
 KW Signal; Glycoprotein; Transmembrane; EGF-like domain;  
 KW Alternative splicing; 23  
 FT SIGNAL 1 23 POTENTIAL.  
 FT PROPEP 24 225 BY SIMILARITY.  
 FT CHAIN 226 769 ADAM 11.  
 FT DOMAIN 226 734 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 735 755 POTENTIAL.  
 FT DOMAIN 756 769 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 226 438 METALLOPROTEASE-LIKE.  
 FT DOMAIN 444 531 DISINTEGRIN-LIKE.  
 FT DOMAIN 532 676 CTS-RICH.  
 FT DOMAIN 677 709 EGF-LIKE.  
 FT DISULEID 349 433 BY SIMILARITY.  
 FT DISULEID 503 516 POTENTIAL.  
 FT DISULEID 677 692 BY SIMILARITY.  
 FT DISULEID 686 698 BY SIMILARITY.  
 FT DISULEID 700 709 BY SIMILARITY.  
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 605 605 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 673 673 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPPLIC 1 99 MISSING (IN SHORT ISOFORM).  
 FT VARSPPLIC 100 104 DIEIN -> MCWLS (IN SHORT ISOFORM).  
 FT VARSPPLIC 595 623 DYICGLGLCVNINSGAPRLCDIYGDISSVT -> PQQGRAV  
 FT VARSPPLIC 624 769 LPELCOHLMSSARGEGGRHQ (IN SHORT ISOFORM).  
 FT CONFLICT 106 106 Q -> H (IN REF. 1).  
 FT CONFLICT 325 325 D -> N (IN REF. 2).  
 SQ SEQUENCE 769 AA: 83409 MM: 59B9C467B6DDF86 CRC64:

Query Match 8.0%; Score 99; DB 1; Length 769;  
 Best Local Similarity 21.4%; Pred. No. 0.17;  
 Matches 55; Conservative 32; Mismatches 90; Indels 80; Gaps 14;

QY 31 LPEICT-----QCPGSYONLSKVAFYCK-----TTRR-----LMLHA---RCC-- 65  
 DB 512 IAECTGDSGCCPNLKLKD--GYCCHGEGRCYGGCKTRDQCOVLMGHAADFFCYE 569  
 QY 66 -LNKGT-----ILGLDLNCSLEDPGPNFQAHTTVIIDLQANPLKGLDANFEGFTQ 118  
 DB 570 KLNVEGTGRSGCGKRGSGWOCCKODVLCF-----LLCVNIGAPRLGLVDISSVTF 624  
 QY 119 LQTLIPQHVNCPEG--INAWNTTTSYID-----NQCQGGKNTL-----CNNTG 160

DB 625 YHQ--GKELDRGCHQVQLADSDLSYEDGTACGPNMCLDHRCLPASAFNFTCPGSG 681  
 QY 161 DPENCPEGSCVPDPGLLOVCADGFGYKC-----MRQSFSLMWF 204  
 DB 682 ERRICSHHGVCNSNG-----KCICPDWTKGDCSHNPLPTSPPTGETERYKGPSTNTI 737  
 QY 205 GLIGATITLSVILLMAT 221  
 DB 738 GSIAQAVLVAIVLGGT 754  
 RESULT 10  
 ID AD11\_MOUSE STANDARD; PRT; 773 AA.  
 AC Q9RIV4; 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ADAM 11 PRECURSOR (A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 11)  
 DE (METALLOPROTEINASE-LIKE, DISINTEGRIN-LIKE, AND CYSTEINE-RICH PROTEIN)  
 DE (MOC).  
 GN ADAM11 OR MDC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=93965303; Pubmed=10433968;  
 RA Sagane K., Yamazaki K., Mizui Y., Tanaka I.;  
 RT "Cloning and chromosomal mapping of mouse ADAM11, ADAM22 and ADAM23.";  
 RL Gene 236:79-86(1999).  
 CC -1- FUNCTION: PROBABLE LIGAND FOR INTEGRIN IN THE BRAIN. THIS IS A NON  
 CC CATALYTIC METALLOPROTEASE-LIKE PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE BRAIN. WEAKLY  
 CC DETECTED IN THE HEART, LIVER AND TESTIS.  
 CC -1- DOMAIN: A CONSERVED MOTIF (AVN(E/D)CD) WITHIN THE DISINTEGRIN-LIKE  
 CC DOMAIN COULD BE INVOLVED IN THE BINDING TO THE INTEGRIN RECEPTOR.  
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC  
 CC METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).





```

FT DOMAIN 714 749 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 751 787 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 789 825 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 827 865 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 867 903 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 905 941 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 943 979 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 981 1017 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1019 1055 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1057 1093 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1095 1141 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1143 1179 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1181 1217 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1219 1263 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1265 1303 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1305 1344 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1346 1382 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1385 1423 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 1446 1486 LIN/NORCH 1.
FT REPEAT 1487 1520 LIN/NORCH 2.
FT REPEAT 1521 1561 LIN/NORCH 3.
FT REPEAT 1567 1590 ANK 1.
FT REPEAT 1595 1644 ANK 2.
FT REPEAT 1648 1698 ANK 3.
FT REPEAT 1698 1748 ANK 4.
FT REPEAT 1748 1804 ANK 5.
FT REPEAT 1804 1860 ANK 6.
FT REPEAT 1860 1910 POLY-GLN (OPA-REPEAT).
FT REPEAT 1910 1944 BY SIMILARITY.
FT REPEAT 1948 1978 BY SIMILARITY.
FT REPEAT 1978 2011 BY SIMILARITY.
FT REPEAT 2011 2044 BY SIMILARITY.
FT REPEAT 2044 2077 BY SIMILARITY.
FT REPEAT 2077 2276 BY SIMILARITY.
FT DOMAIN 2265 2276 BY SIMILARITY.
FT DISULFID 25 35 BY SIMILARITY.
FT DISULFID 29 45 BY SIMILARITY.
FT DISULFID 47 56 BY SIMILARITY.
FT DISULFID 62 73 BY SIMILARITY.
FT DISULFID 67 86 BY SIMILARITY.
FT DISULFID 88 97 BY SIMILARITY.
FT DISULFID 105 116 BY SIMILARITY.
FT DISULFID 110 126 BY SIMILARITY.
FT DISULFID 128 137 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 163 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 181 194 BY SIMILARITY.
FT DISULFID 194 203 BY SIMILARITY.
FT DISULFID 188 205 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY.
FT DISULFID 226 242 BY SIMILARITY.
FT DISULFID 244 253 BY SIMILARITY.
FT DISULFID 260 271 BY SIMILARITY.
FT DISULFID 271 280 BY SIMILARITY.
FT DISULFID 265 280 BY SIMILARITY.
FT DISULFID 282 291 BY SIMILARITY.
FT DISULFID 291 311 BY SIMILARITY.
FT DISULFID 298 311 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 338 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 375 386 BY SIMILARITY.
FT DISULFID 397 408 BY SIMILARITY.
FT DISULFID 380 397 BY SIMILARITY.
FT DISULFID 399 408 BY SIMILARITY.
FT DISULFID 415 428 BY SIMILARITY.
FT DISULFID 422 437 BY SIMILARITY.
FT DISULFID 437 448 BY SIMILARITY.
FT DISULFID 448 455 BY SIMILARITY.
FT DISULFID 455 466 BY SIMILARITY.
FT DISULFID 460 475 BY SIMILARITY.
FT DISULFID 477 486 BY SIMILARITY.
FT DISULFID 493 503 BY SIMILARITY.
FT DISULFID 498 512 BY SIMILARITY.
FT DISULFID 512 523 BY SIMILARITY.
FT DISULFID 523 541 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 541 550 BY SIMILARITY.
FT DISULFID 550 561 BY SIMILARITY.
FT DISULFID 561 578 BY SIMILARITY.
FT DISULFID 578 587 BY SIMILARITY.
FT DISULFID 587 598 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.

```

```

FT DISULFID 605 616 BY SIMILARITY.
FT DISULFID 610 625 BY SIMILARITY.
FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 648 662 BY SIMILARITY.
FT DISULFID 664 673 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT DISULFID 691 700 BY SIMILARITY.
FT DISULFID 702 711 BY SIMILARITY.
FT DISULFID 711 728 BY SIMILARITY.
FT DISULFID 728 737 BY SIMILARITY.
FT DISULFID 737 748 BY SIMILARITY.
FT DISULFID 748 755 BY SIMILARITY.
FT DISULFID 755 766 BY SIMILARITY.
FT DISULFID 766 775 BY SIMILARITY.
FT DISULFID 777 786 BY SIMILARITY.
FT DISULFID 793 804 BY SIMILARITY.
FT DISULFID 804 813 BY SIMILARITY.
FT DISULFID 813 824 BY SIMILARITY.
FT DISULFID 824 831 BY SIMILARITY.
FT DISULFID 831 842 BY SIMILARITY.
FT DISULFID 842 853 BY SIMILARITY.
FT DISULFID 853 864 BY SIMILARITY.
FT DISULFID 864 871 BY SIMILARITY.
FT DISULFID 871 882 BY SIMILARITY.
FT DISULFID 882 891 BY SIMILARITY.
FT DISULFID 891 902 BY SIMILARITY.
FT DISULFID 902 920 BY SIMILARITY.
FT DISULFID 920 929 BY SIMILARITY.
FT DISULFID 929 940 BY SIMILARITY.
FT DISULFID 940 958 BY SIMILARITY.
FT DISULFID 958 967 BY SIMILARITY.
FT DISULFID 967 978 BY SIMILARITY.
FT DISULFID 978 1023 BY SIMILARITY.
FT DISULFID 1023 1034 BY SIMILARITY.
FT DISULFID 1034 1043 BY SIMILARITY.
FT DISULFID 1043 1054 BY SIMILARITY.

Query Match 7.9% Score 98.5; DB 1; Length 2437;
Best local Similarity 20.6%; Pred. No. 0.65;
Matches 42; Conservative 21; Mismatches 58; Indels 83; Gaps 10;

QY 34 ICTQCPG-----SVQNLKVAFYCKTTRREIMHARCLNQKGT-----ILGIDLQNC 80
DQ 396 ICTCPGYGSAQNDIDCSLQANPC-----HGRCCLMTKSGFOCKLOGEGPRC 448
QY 81 SLEDPGPNHQHTTYITDQLANPLKGLD--NTERGFYQLOTLLPQ----- 126
DQ 449 EMD-----VNECKSNPCQNDATCDOIIGF--HCICMPGEYGFQCOINSDD 492
QY 127 -----HYNCPGGINAMNTTYSYDNOICOGKKNLNNGTDPENCPEP 168
DQ 493 CASQPCINCKIDKINSFHCCEPKGFS-----GSLCOVDYDECAST-----PC-KN 537
QY 169 GSCVPDPGPGILQCCVACADGFHYK 192
DQ 538 GAKCTDGPNNKYCECPRGFSGIMC 561

RESULT 12
ID L112 CAEEL STANDARD; PRT; 1429 AA.
AC P14585;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIN-12 PROTEIN PRECURSOR.
GN LIN-12 OR R107.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=8834747; PubMed=3419531;
RA Yochem J., Weston K., Greenwald I.;

```

RT "The Caenorhabditis elegans lin-12 gene encodes a transmembrane  
 RT protein with overall similarity to Drosophila Notch."  
 RT Nature 335:547-550(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2.  
 RX MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Faveille A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnson L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shonkhen R.,  
 RA Sims M., Smalton N., Smith A., Smith M., Sponhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Watson R., Watson A., Weinstock L., Wilkinson-Spreat J.,  
 RA Woldman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans."  
 RL Nature 368:32-38(1994).  
 CC -1- FUNCTION: LIN-12 IS INVOLVED IN SEVERAL CELL FATES DECISIONS THAT  
 CC REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12  
 CC ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES  
 CC EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: HIGH TO C.ELEGANS GLP-1.  
 CC -1- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 DR EMBL: M12069; AAA70191.1; -;  
 DR EMBL: Z14092; CAA78474.1; -;  
 DR PIR: S06434; S06434.  
 DR HSSP: P00740; 11XA.  
 DR Wormpep: R107.8; CE00274.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF-2.  
 DR InterPro: IPR001881; EGF-Ca.  
 DR InterPro: IPR000800; Notch.  
 DR Pfam: PF00023; ank; 6  
 DR Pfam: PF00008; EGF; 13.  
 DR Pfam: PF00066; notch; 3.  
 DR SMART: SM00248; ANK; 4.  
 DR SMART: SM00179; EGF-Ca; 2.  
 DR SMART: SM00004; NL; 2.  
 DR PROSITE: PSS00088; ANK\_REPEAT; 3.  
 DR PROSITE: PSS0297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PSS0010; ASX\_HYDROXYL; 3.  
 DR PROSITE: PSS0022; EGF\_1; 12.  
 DR PROSITE: PSS0186; EGF\_2; 11.  
 DR PROSITE: PSS0187; EGF-Ca; 2.  
 KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;  
 KW Glycoprotein; Signal;  
 FT SIGNAL 1 15  
 FT CHAIN 16 1429  
 FT DOMAIN 16 908  
 FT TRANSSEM 909 931  
 FT DOMAIN 932 1429  
 FT DOMAIN 20 61  
 FT DOMAIN 114 150  
 FT DOMAIN 152 190  
 EGF-Like 2.  
 EGF-Like 3, CALCIUM-BINDING (POTENTIAL).

FT	DOMAIN	201	246	EGF-Like 4.
FT	DOMAIN	250	285	EGF-Like 5.
FT	DOMAIN	287	323	EGF-Like 6.
FT	DOMAIN	323	363	EGF-Like 7.
FT	DOMAIN	365	402	EGF-Like 8.
FT	DOMAIN	404	441	EGF-Like 9.
FT	DOMAIN	449	492	EGF-Like 10.
FT	DOMAIN	503	541	EGF-Like 11.
FT	DOMAIN	543	579	EGF-Like 12.
FT	DOMAIN	582	619	EGF-Like 13.
FT	REPEAT	635	669	LIN/NOTCH 1.
FT	REPEAT	670	710	LIN/NOTCH 2.
FT	REPEAT	711	750	LIN/NOTCH 3.
FT	REPEAT	1093	1122	ANK 1.
FT	REPEAT	1126	1158	ANK 2.
FT	REPEAT	1162	1194	ANK 3.
FT	REPEAT	1206	1236	ANK 4.
FT	REPEAT	1240	1269	ANK 5.
FT	DISULFID	24	35	BY SIMILARITY.
FT	DISULFID	29	49	BY SIMILARITY.
FT	DISULFID	51	60	BY SIMILARITY.
FT	DISULFID	118	129	BY SIMILARITY.
FT	DISULFID	123	138	BY SIMILARITY.
FT	DISULFID	140	149	BY SIMILARITY.
FT	DISULFID	156	169	BY SIMILARITY.
FT	DISULFID	163	178	BY SIMILARITY.
FT	DISULFID	180	189	BY SIMILARITY.
FT	DISULFID	205	227	BY SIMILARITY.
FT	DISULFID	221	234	BY SIMILARITY.
FT	DISULFID	236	245	BY SIMILARITY.
FT	DISULFID	254	264	BY SIMILARITY.
FT	DISULFID	259	273	BY SIMILARITY.
FT	DISULFID	275	284	BY SIMILARITY.
FT	DISULFID	291	302	BY SIMILARITY.
FT	DISULFID	296	311	BY SIMILARITY.
FT	DISULFID	313	322	BY SIMILARITY.
FT	DISULFID	327	339	BY SIMILARITY.
FT	DISULFID	334	351	BY SIMILARITY.
FT	DISULFID	353	362	BY SIMILARITY.
FT	DISULFID	369	381	BY SIMILARITY.
FT	DISULFID	375	390	BY SIMILARITY.
FT	DISULFID	392	401	BY SIMILARITY.
FT	DISULFID	408	419	BY SIMILARITY.
FT	DISULFID	413	429	BY SIMILARITY.
FT	DISULFID	431	440	BY SIMILARITY.
FT	DISULFID	507	518	BY SIMILARITY.
FT	DISULFID	512	529	BY SIMILARITY.
FT	DISULFID	531	540	BY SIMILARITY.
FT	DISULFID	547	558	BY SIMILARITY.
FT	DISULFID	552	567	BY SIMILARITY.
FT	DISULFID	569	578	BY SIMILARITY.
FT	DISULFID	586	597	BY SIMILARITY.
FT	DISULFID	591	607	BY SIMILARITY.
FT	DISULFID	609	618	BY SIMILARITY.
FT	CARBOHYD	41	41	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	165	165	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	194	194	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	378	378	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	515	515	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	623	623	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	751	751	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	754	754	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	900	900	N-LINKED (GLCNAc. . .) (POTENTIAL).
SEQ	SEQUENCE	1429 AA;	157115 MW;	255EDV/A62C025DB CRC64;

Query Match  
 Best Local Similarity 7.9%; Score 98; DB 1; Length 1429;  
 Matches 42; Conservative 27; Mismatches 57; Indels 46; Gaps 12;  
 34 ICTGCGSYONLSKVAFA--YCKTRREIMLHARC-----CINOKTIIIGLD--LQNSLE 83  
 Db 263 VCISFGGQCICIDGQISGYSIQBGKDMCVNKKCBAGSKCTN-----GVNSYFCDCPE 316

```

OY 84 DPGPHQAHHTVILIDLANPLKGLANTFRGOTLOTLLPOHVC---PGGINAMNTI 140
DB 317 RTGPCYCK-----MDCSNIP---DICH--HG--TCIDPISKEAFECOCERG----- 356
OY 141 TSYINQICOGQKNCNNTGDEPCENGSCVPDGPGLLOCCVACADPHGYNC 192
DB 357 ---YEGILCEQKNECLS---ENMCLNNGTCV-NLPGSFRCDCARGFGKWC 401

RESULT 13
ID NTCL_MOUSE STANDARD: PRT: 2531 AA.
AC 001705:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (NOTCH PROTEIN).
GN NOTCH1 OR NOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=93194170; PubMed=8449489;
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
RT Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT homologue of Drosophila Notch-1.
RL Genomics 15:259-264(1993).
RN [2]
RP SEQUENCE OF 1551-2170 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=93048835; PubMed=1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA Greenspan R.J., McMahon A.P., Gridley T.
RT Expression pattern of Notch, a mouse homologue of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
RT development.
RT Development 115:737-744(1992).
RL Development 115:737-744(1992).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z11886; CAAT7941.1; -.
DR HSSP: P00740; 1IXA.
DR MGD: MGI:97363; Notch1.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; ASX_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001481; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000800; Notch.
DR Pfam: PF000023; ank. 6.
DR Pfam: PF000068; EGF_35.
DR Pfam: PF000066; notch. 3.
DR PRINTS: PR00010; EGFBL00D.
DR SMART: SM00248; ANK. 3.
DR SMART: SM00179; EGF_CA. 23.
DR SMART: SM00001; EGF-like. 13.
DR SMART: SM00004; NL. 2.
DR PROSITE: PS50088; ANK_REPEAT. 2.
DR PROSITE: PS50297; ANK_REGION. 1.
DR PROSITE: PS00010; ASX_HYDROXYL. 22.
DR PROSITE: PS00022; EGF_1. 34.
DR PROSITE: PS01186; EGF_2. 27.
DR PROSITE: PS01187; EGF_Ca. 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 2531
FT DOMAIN 19 1725
FT TRANSMEM 1726 1746
FT TRANSMEM 1747 2531
FT DOMAIN 20 58
FT DOMAIN 59 99
FT DOMAIN 102 139
FT DOMAIN 140 176
FT DOMAIN 178 216
FT DOMAIN 218 255
FT DOMAIN 257 293
FT DOMAIN 295 333
FT DOMAIN 335 371
FT DOMAIN 372 410
FT DOMAIN 412 450
FT DOMAIN 452 488
FT DOMAIN 490 526
FT DOMAIN 528 564
FT DOMAIN 566 601
FT DOMAIN 603 639
FT DOMAIN 641 676
FT DOMAIN 678 714
FT DOMAIN 716 751
FT DOMAIN 753 789
FT DOMAIN 791 827
FT DOMAIN 829 867
FT DOMAIN 869 905
FT DOMAIN 907 943
FT DOMAIN 945 981
FT DOMAIN 983 1019
FT DOMAIN 1021 1057
FT DOMAIN 1059 1095
FT DOMAIN 1097 1143
FT DOMAIN 1145 1181
FT DOMAIN 1183 1219
FT DOMAIN 1221 1265
FT DOMAIN 1267 1305
FT DOMAIN 1307 1346
FT DOMAIN 1348 1384
FT DOMAIN 1387 1426
FT DOMAIN 1445 1480
FT DOMAIN 1449 1486
FT REPEAT 1481 1522
FT REPEAT 1523 1562
FT REPEAT 1562 1597
FT REPEAT 1597 1642
FT REPEAT 1642 1687
FT REPEAT 1687 1732
FT REPEAT 1732 1777
FT REPEAT 1777 1822
FT REPEAT 1822 1867
FT DISULFID 24 37
FT DISULFID 31 46
FT DISULFID 63 74
FT DISULFID 68 87
FT DISULFID 89 98
FT DISULFID 106 117
FT DISULFID 111 127
FT DISULFID 129 138
FT DISULFID 144 155
FT DISULFID 149 164
FT DISULFID 166 175
FT DISULFID 182 195
FT DISULFID 189 204

```

FT DISULFID 206 215 BY SIMILARITY.  
 FT DISULFID 222 233 BY SIMILARITY.  
 FT DISULFID 227 243 BY SIMILARITY.  
 FT DISULFID 245 254 BY SIMILARITY.  
 FT DISULFID 261 272 BY SIMILARITY.  
 FT DISULFID 266 281 BY SIMILARITY.  
 FT DISULFID 283 292 BY SIMILARITY.  
 FT DISULFID 289 312 BY SIMILARITY.  
 FT DISULFID 306 321 BY SIMILARITY.  
 FT DISULFID 323 332 BY SIMILARITY.  
 FT DISULFID 339 350 BY SIMILARITY.  
 FT DISULFID 344 359 BY SIMILARITY.  
 FT DISULFID 361 370 BY SIMILARITY.  
 FT DISULFID 376 387 BY SIMILARITY.  
 FT DISULFID 381 398 BY SIMILARITY.  
 FT DISULFID 400 409 BY SIMILARITY.  
 FT DISULFID 416 429 BY SIMILARITY.  
 FT DISULFID 423 438 BY SIMILARITY.  
 FT DISULFID 440 449 BY SIMILARITY.  
 FT DISULFID 456 467 BY SIMILARITY.  
 FT DISULFID 478 487 BY SIMILARITY.  
 FT DISULFID 494 505 BY SIMILARITY.  
 FT DISULFID 514 525 BY SIMILARITY.  
 FT DISULFID 532 543 BY SIMILARITY.  
 FT DISULFID 537 552 BY SIMILARITY.  
 FT DISULFID 554 563 BY SIMILARITY.  
 FT DISULFID 570 580 BY SIMILARITY.  
 FT DISULFID 575 589 BY SIMILARITY.  
 FT DISULFID 591 600 BY SIMILARITY.  
 FT DISULFID 607 618 BY SIMILARITY.  
 FT DISULFID 612 627 BY SIMILARITY.  
 FT DISULFID 629 638 BY SIMILARITY.  
 FT DISULFID 645 655 BY SIMILARITY.  
 FT DISULFID 660 664 BY SIMILARITY.  
 FT DISULFID 666 675 BY SIMILARITY.  
 FT DISULFID 682 693 BY SIMILARITY.  
 FT DISULFID 687 702 BY SIMILARITY.  
 FT DISULFID 704 713 BY SIMILARITY.  
 FT DISULFID 720 730 BY SIMILARITY.  
 FT DISULFID 725 739 BY SIMILARITY.  
 FT DISULFID 741 750 BY SIMILARITY.  
 FT DISULFID 757 768 BY SIMILARITY.  
 FT DISULFID 762 777 BY SIMILARITY.  
 FT DISULFID 779 788 BY SIMILARITY.  
 FT DISULFID 795 806 BY SIMILARITY.  
 FT DISULFID 800 815 BY SIMILARITY.  
 FT DISULFID 817 826 BY SIMILARITY.  
 FT DISULFID 833 844 BY SIMILARITY.  
 FT DISULFID 838 855 BY SIMILARITY.  
 FT DISULFID 857 866 BY SIMILARITY.  
 FT DISULFID 873 884 BY SIMILARITY.  
 FT DISULFID 893 893 BY SIMILARITY.  
 FT DISULFID 904 922 BY SIMILARITY.  
 FT DISULFID 911 922 BY SIMILARITY.  
 FT DISULFID 931 942 BY SIMILARITY.  
 FT DISULFID 933 942 BY SIMILARITY.  
 FT DISULFID 967 998 BY SIMILARITY.  
 FT DISULFID 992 1007 BY SIMILARITY.  
 FT DISULFID 1009 1018 BY SIMILARITY.  
 FT DISULFID 1025 1036 BY SIMILARITY.  
 FT DISULFID 1030 1045 BY SIMILARITY.  
 FT DISULFID 1047 1056 BY SIMILARITY.  
 FT DISULFID 1063 1074 BY SIMILARITY.  
 FT DISULFID 1068 1083 BY SIMILARITY.

Query Match 7.9% Score 98; DB 1; Length 2531;  
 Best Local Similarity 22.6% Pred. No. 0.76;

Matches 44; Conservative 16; Mismatches 71; Indels 64; Gaps 10;

QY 34 ICTGCGSVQVN--LSKVAFYCKTTRRLMLHRCCLNCKGTTGLDLCNCSLEDPGRNFHQ 91  
 ||| ||| | : | ||| : | : | : |

Db 397 ICT-CPSSYTGPGACSDVDECDLGANRCEHAKCLNTLIGSTECOCLOGYT---GPCE- 450  
 QY 92 AHRTYVLIIDQ---ANPLKGLDANFRGPTOLQTLIPQH----- 127  
 Db 451 -----IDVNECISNPKQND-ATCLDQIGERQCIOMGVEYVCEINTDECASSPLCHNG 503  
 QY 128 -----VACPGIAHANNITTSYIDNQCQKNCNNTGDPKCPKNGSCVPDPG 177  
 Db 504 HCMKXIHFFQCCPGFVN-----GHLCQYDVDECAST-----PKRNGAKCL-DGPN 548  
 QY 178 ILQVCVADGFHGYKC 192  
 Db 549 TYTCVCTEGYTGTHC 563

RESULT 14  
 COMP\_RAT  
 ID COMP\_RAT STANDARD; PRT; 755 AA.  
 AC P35444;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE CARTILAGE OLIGOMERIC MATRIX PROTEIN PRECURSOR (COMP).  
 GN COMP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cartilage;  
 RX MEDLINE=93054522; PubMed=1429587;  
 RA Oldberg A., Antonsson P., Lindblom K., Heinegaard D.;  
 RT "COMP (cartilage oligomeric matrix protein) is structurally related  
 to the thrombospondins.";  
 RL J. Biol. Chem. 267:22346-22350(1992).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 27-72.  
 RX MEDLINE=97020114; PubMed=8864111;  
 RA Malashkevich V.N., Kammerer R.A., Efimov V.P., Schulthess T.,  
 RA Engel J.;  
 RT "The crystal structure of a five-stranded coiled coil in COMP: a  
 prototype ion channel?";  
 RL Science 274:761-765(1996).  
 CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X72914; CAA51419.1; -;  
 DR PIR: A44315; A44315.  
 DR PDB: 1VDF; 08-OCT-97.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR00367; tsp\_3.  
 DR Pfam: PF02412; tsp\_3; 9.  
 DR SMART: SM00179; EGF\_CA; 2.  
 DR SMART: SM00001; EGF-like; 2.  
 DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF\_CA; 2.  
 DR GlycoProtein: Cell adhesion; Calcium-binding; Repeat; EGF-like domain;  
 KW Signal; 3D-structure.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 755 CARTILAGE OLIGOMERIC MATRIX PROTEIN.

FT DOMAIN 21 84 N-TERMINAL.  
 FT DOMAIN 85 124 EGF-LIKE 1.  
 FT DOMAIN 125 177 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 178 220 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 223 265 EGF-LIKE 4.  
 FT DOMAIN 293 330 TSP TYPE-3 1.  
 FT DOMAIN 331 353 TSP TYPE-3 2.  
 FT DOMAIN 354 389 TSP TYPE-3 3.  
 FT DOMAIN 390 412 TSP TYPE-3 4.  
 FT DOMAIN 413 450 TSP TYPE-3 5.  
 FT DOMAIN 451 486 TSP TYPE-3 6.  
 FT DOMAIN 487 522 TSP TYPE-3 7.  
 FT DOMAIN 523 755 C-TERMINAL.  
 FT DISULFID 68 68 INTERCHAIN (PROBABLE).  
 FT DISULFID 71 71 INTERCHAIN (PROBABLE).  
 FT DISULFID 89 100 BY SIMILARITY.  
 FT DISULFID 94 109 BY SIMILARITY.  
 FT DISULFID 112 123 BY SIMILARITY.  
 FT DISULFID 129 140 BY SIMILARITY.  
 FT DISULFID 134 149 BY SIMILARITY.  
 FT DISULFID 152 176 BY SIMILARITY.  
 FT DISULFID 182 195 BY SIMILARITY.  
 FT DISULFID 189 204 BY SIMILARITY.  
 FT DISULFID 207 219 BY SIMILARITY.  
 FT DISULFID 227 241 BY SIMILARITY.  
 FT DISULFID 233 251 BY SIMILARITY.  
 FT DISULFID 253 264 BY SIMILARITY.  
 FT CARBOHYD 119 119 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 740 740 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 755 AA; 82663 MW; AB48888FE03C598 CRC64;

Query Match 7 8%; Score 97.5; DB 1; Length 755;  
 Best Local Similarity 20.6%; Pred. No. 0.22;  
 Matches 55; Conservative 24; Mismatches 87; Indels 101; Gaps 10;

QY 14 PMAAALLALGVERALA--LPEICTQCGSVONISKVAFYCKTTRMLH----- 61  
 DB 3 PTCACVLAALRATGOGGQIPGLGDLAPQMLRELOETMAALADVRELLRHVRKEITFLK 62  
 QY 62 -----ARCLINOKGTLLGLDLONCSLEDDGPNP-----HOAH 93  
 DB 63 NTWECDACGMPARTPGSVFVALCAPGSCFPVVCETATGARGCPRPYTGNSH 122  
 QY 94 TTVIIDLQANPL--KGDLANTEFRTQLOTLILPOHVN--CPEGIN-----AMNTTSTID 145  
 DB 123 CTDVNEGNAPRCPPRVACINTSPGF-----HCEACPPGFGSPTHGVLFAKTN 172  
 QY 146 NOIC-----OGKN-----LCNMT-----GDP 162  
 DB 173 KQVCTDINECEGQHNCPVNSVCNTRGSGFCGCPGPGVGDORSQCORRQGHPCPDSP 232  
 QY 163 EMCPEGSCVDPGFLGLQVCADGPHG 189  
 DB 233 SPCHKADCLIERDGSRSVCVAGWAG 259

RESULT 15  
 TGFβ\_RAT STANDARD; PRT; 1712 AA.  
 AC 000918;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE LATENT TRANSFORMING GROWTH FACTOR BETA BINDING PROTEIN 1 PRECURSOR  
 DE (TRANSFORMING GROWTH FACTOR BETA-1 BINDING PROTEIN 1) (TGF-βET-1-BP-  
 DE 1) (TRANSFORMING GROWTH FACTOR BETA-1 MASKING PROTEIN, LARGE SUBUNIT).  
 GN LTPPL.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 ON NCBI\_TaxID=10116;  
 RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=91062373; PubMed=247454;  
 RA Tsuji T., Okada F., Yamaguchi K., Nakamura T.;  
 RT "Molecular cloning of the large subunit of transforming growth factor  
 type beta masking protein and expression of the mRNA in various rat  
 tissues.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8835-8839(1990).  
 CC -1- SUBUNIT: THE LARGE LATENT COMPLEX OF TGF-βET-1 FROM PLATELETS IS  
 CC COMPOSED OF THE TGF-βET-1 MOLECULE NONCOVALENTLY ASSOCIATED WITH  
 CC A MASKING PROTEIN CONSISTING OF A DISULFIDE-BONDED COMPLEX OF A  
 CC DIMER OF THE N-TERMINAL PROPEPTIDE OF THE TGF-βET-1 PRECURSOR AND  
 CC A THIRD COMPONENT DENOTED TGF-βET-1-BP (OR MP LARGE SUBUNIT).  
 CC TGF-βET-1-BP DOES NOT BIND DIRECTLY TO ACTIVE TGF-βET-1.  
 CC -1- SIMILARITY: CONTAINS 18 EGF-LIKE DOMAINS.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M55431; AAA42235.1; -  
 PIR: A38261; A38261.

DR HSSP; P16109; 1FSB.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR00212; Tβ.  
 DR Pfam; PF00008; EGF; 16.  
 DR Pfam; PF00683; Tβ; 4.  
 DR SMART; SM00179; EGF\_Ca; 13.  
 DR SMART; SM00001; EGF-like; 5.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 13.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 10.  
 DR PROSITE; PS01187; EGF\_Ca; 15.  
 DR Growth factor binding; Repeat; 20  
 FT SIGNAL 1 20  
 FT PROPEP 21 736  
 FT CHAIN 737 1577  
 FT SITE 734 736  
 FT SITE 1575 1577  
 FT PROPEP 1578 1712  
 FT DOMAIN 181 213  
 FT DOMAIN 391 423  
 FT REPEAT 551 604  
 FT REPEAT 618 658  
 FT DOMAIN 671 721  
 FT REPEAT 721 721  
 FT DOMAIN 906 906  
 FT DOMAIN 907 948  
 FT DOMAIN 949 989  
 FT DOMAIN 990 1029  
 FT DOMAIN 1030 1070  
 FT DOMAIN 1071 1111  
 FT DOMAIN 1112 1152  
 FT DOMAIN 1153 1193  
 FT DOMAIN 1194 1235  
 FT DOMAIN 1236 1277  
 FT DOMAIN 1278 1320  
 FT DOMAIN 1340 1392  
 FT REPEAT 1415 1457  
 FT DOMAIN 1458 1498  
 FT REPEAT 1517 1562  
 FT DOMAIN 1612 1652  
 FT DOMAIN 1653 1697  
 FT DISULFID 185 195  
 FT DISULFID 189 201  
 FT DISULFID 203 212  
 FT DISULFID 395 403  
 FT DISULFID 411

LATENT TRANSFORMING GROWTH FACTOR BETA  
 BINDING PROTEIN 1.  
 CLEAVAGE (POTENTIAL).  
 POTENTIAL.  
 EGF-LIKE 1.  
 EGF-LIKE 2.  
 INTERNAL REPEAT 1.  
 EGF-LIKE 3. CALCIUM-BINDING (POTENTIAL).  
 INTERNAL REPEAT 2.  
 EGF-LIKE 4. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 5. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 6. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 7. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 8. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 9. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 10. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 14. CALCIUM-BINDING (POTENTIAL).  
 INTERNAL REPEAT 3.  
 EGF-LIKE 15. CALCIUM-BINDING (POTENTIAL).  
 EGF-LIKE 16. CALCIUM-BINDING (POTENTIAL).  
 INTERNAL REPEAT 4.  
 EGF-LIKE 17.  
 EGF-LIKE 18. CALCIUM-BINDING (POTENTIAL).  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.  
 BY SIMILARITY.

FT DISULFID 413 422 BY SIMILARITY.  
FT DISULFID 622 633 BY SIMILARITY.  
FT DISULFID 628 642 BY SIMILARITY.  
FT DISULFID 644 657 BY SIMILARITY.  
FT DISULFID 869 881 BY SIMILARITY.  
FT DISULFID 876 890 BY SIMILARITY.  
FT DISULFID 892 905 BY SIMILARITY.  
FT DISULFID 911 923 BY SIMILARITY.  
FT DISULFID 918 932 BY SIMILARITY.  
FT DISULFID 934 947 BY SIMILARITY.  
FT DISULFID 953 964 BY SIMILARITY.  
FT DISULFID 959 973 BY SIMILARITY.  
FT DISULFID 976 988 BY SIMILARITY.  
FT DISULFID 994 1005 BY SIMILARITY.  
FT DISULFID 1000 1014 BY SIMILARITY.  
FT DISULFID 1017 1028 BY SIMILARITY.  
FT DISULFID 1034 1045 BY SIMILARITY.  
FT DISULFID 1040 1054 BY SIMILARITY.  
FT DISULFID 1056 1069 BY SIMILARITY.  
FT DISULFID 1075 1086 BY SIMILARITY.  
FT DISULFID 1097 1110 BY SIMILARITY.  
FT DISULFID 1116 1127 BY SIMILARITY.  
FT DISULFID 1122 1136 BY SIMILARITY.  
FT DISULFID 1138 1151 BY SIMILARITY.  
FT DISULFID 1157 1169 BY SIMILARITY.  
FT DISULFID 1164 1178 BY SIMILARITY.  
FT DISULFID 1180 1192 BY SIMILARITY.  
FT DISULFID 1198 1210 BY SIMILARITY.  
FT DISULFID 1204 1219 BY SIMILARITY.  
FT DISULFID 1221 1234 BY SIMILARITY.  
FT DISULFID 1240 1252 BY SIMILARITY.  
FT DISULFID 1246 1261 BY SIMILARITY.  
FT DISULFID 1263 1276 BY SIMILARITY.  
FT DISULFID 1282 1294 BY SIMILARITY.  
FT DISULFID 1289 1303 BY SIMILARITY.  
FT DISULFID 1305 1319 BY SIMILARITY.  
FT DISULFID 1419 1432 BY SIMILARITY.  
FT DISULFID 1427 1441 BY SIMILARITY.  
FT DISULFID 1443 1456 BY SIMILARITY.  
FT DISULFID 1462 1473 BY SIMILARITY.  
FT DISULFID 1468 1482 BY SIMILARITY.  
FT DISULFID 1484 1497 BY SIMILARITY.  
FT DISULFID 1616 1627 BY SIMILARITY.  
FT DISULFID 1622 1636 BY SIMILARITY.  
FT DISULFID 1638 1651 BY SIMILARITY.  
FT DISULFID 1657 1672 BY SIMILARITY.  
FT DISULFID 1667 1681 BY SIMILARITY.  
FT DISULFID 1683 1696 BY SIMILARITY.  
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 416 416 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1042 1042 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1242 1242 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1357 1357 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1712 AA; 186598 MW; 650BCA691FD134 CRC64;

Query Match 7.8%; Score 97.5; DB 1; Length 1712;  
Best Local Similarity 23.5%; Pred. No. 0.55;  
Matches 50; Conservative 16; Mismatches 64; Indels 83; Gaps 13;

OY 35 CTQCPGVSQVNSKVAFYKTT-----RELMLHARC-----CLNKG---TILG---L 75  
DB 1095 CT-C-GGGYOLSAARDQCEHDECEHRLCSHGCRNTEGSGFCLCNGRYASVIGDHC 1152  
OY 76 DLQNSLEDPPGNFNGHATTVIIDDQANPLKGLDANTFRGFTOLTLPLQHVNCPGGIN 135  
DB 1153 DINEC-LBD-----SSVCGGDCINTAGSY-----DCTCPDGL- 1184  
OY 136 AMNTTSTYIDNQICGQKNCNNTGDPMPENSGSCVDPGGLQVCADGF----- 187

Db 1185 -----QLDNKNGCOD----INECAQPGICAPHGECL-NTQGSFHVCQEGSFISADGRT 1233  
OY 188 -----HGKCMRQGSFSLIMFPG 205  
DB 1234 CEDIDECVNNTVCDSHGFCDNNTAGSFRCICYOG 1266

Search completed: March 28, 2002, 09:14:46  
Job time: 875 sec

Thu Mar 28 09:21:34 2002

us-09-726-348-2\_1.rsp

---

Page 22



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:16:42 ; Search time 21.01 Seconds  
(without alignments)  
399.631 Million cell updates/sec

Title: US-09-726-348-2

Sequence: 1 MAPHGPSLITLVPMMAALL.....TTLSVSTILWATQRRKAKTS 229

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1802

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database: SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5	2.2	15	1	ASPL_LACSN
2	5	2.2	25	1	ALR_PSEFL
3	5	2.2	28	1	ACON_CANAL
4	5	2.2	29	1	HS98_NEUCR
5	5	2.2	30	1	VAAL_EQUAR
6	4	1.7	9	1	LITR_PHYRO
7	4	1.7	10	1	TKL4_LOCM1
8	4	1.7	10	1	TKSL_AEDAE
9	4	1.7	10	1	TRP8_LEUMA
10	4	1.7	12	1	FREL_LITIN
11	4	1.7	12	1	PORD_METTM
12	4	1.7	12	1	V25K_WSSV
13	4	1.7	13	1	CHEP_PARID
14	4	1.7	15	1	KLOM_LUMTE
15	4	1.7	15	1	BKHM_PHYPA
16	4	1.7	16	1	CRRB_RAT
17	4	1.7	16	1	EX_RAT
18	4	1.7	16	1	RIPK_TRIKI
19	4	1.7	17	1	BOL5_MEGBE
20	4	1.7	17	1	SRY_URSAR
21	4	1.7	18	1	DRPH_UCAPU
22	4	1.7	18	1	TKN2_SCYCA
23	4	1.7	19	1	HMD_METWO
24	4	1.7	19	1	ITHA_PERAM
25	4	1.7	19	1	MDH_SHEPU
26	4	1.7	20	1	CPB5_PAPSP
27	4	1.7	20	1	CPB3_CAVPO
28	4	1.7	20	1	MTF_PIG
29	4	1.7	20	1	PYRX_PSEFL
30	4	1.7	21	1	EPAB_HUMAN
31	4	1.7	21	1	MDH_KLEPN
32	4	1.7	22	1	ATP6_COTUA
33	4	1.7	22	1	LPT_SERMA

## ALIGNMENTS

34	4	1.7	23	1	CR41_LITCE	P56242 litoria cae
35	4	1.7	23	1	CR42_LITCE	P56243 litoria cae
36	4	1.7	23	1	CR43_LITCE	P56244 litoria cae
37	4	1.7	23	1	GP50_BPSP1	O48404 bacterioph
38	4	1.7	23	1	PRP1_RAT	P10165 rattus norv
39	4	1.7	24	1	IPER_STRPR	P45440 streptomyc
40	4	1.7	24	1	PSAC_CUCSA	P42046 cucumis sat
41	4	1.7	24	1	RAN_XENLA	P52301 xenopus lae
42	4	1.7	25	1	ACP_ERYLO	P80919 erythroba
43	4	1.7	25	1	ATP0_SPIOL	P80082 spinnacia ol
44	4	1.7	25	1	ATP0_MICLU	P80285 micrococ
45	4	1.7	25	1	COXO_ONCMY	P80334 oncornynch

RESULT 1  
ID ASPL\_LACSN STANDARD: PRT: 15 AA.

AC P82648: 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DE 20-AUG-2001 (Rel. 40, Last annotation update)  
DE ACID SHOCK PROTEIN 1 (FRAGMENT)  
OS Lactobacillus sanfranciscensis (Lactobacillus sanfrancisco).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1625;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=CRI.  
RX PubMed=11429463:  
RA De Angelis M., Bini L., Pallini V., Cocconcelli P.S., Gobetti M.;  
RT "The acid-stress response in Lactobacillus sanfranciscensis CRL 1";  
RL Microbiology 147:1863-1873(2001).  
CC -I- INDUCTION: OVEREXPRESSED IN ACID ENVIRONMENTS.  
FT NON TER 15  
SO SEQUENCE 15 AA; 1509 MW; 575853B4DFB030A8 CRC64;

Query Match 2.2%; Score 5; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 26+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 GTILG 74  
Db 9 GTILG 74

RESULT 2  
ID ALR\_PSEFL STANDARD: PRT: 25 AA.

DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DE 01-FEB-1994 (Rel. 28, Last annotation update)  
DE ALANINE RACEMASE (EC 5.1.1.1) (FRAGMENT).  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=294;  
RN [1]  
RP SEQUENCE.  
RC STRAIN=TKS-2;  
RX MEDLINE=93169026; PubMed=7763424;  
RA Yokoyama K., Kawai H., Endo K., Iim Y.H., Esaki N., Soda K.;  
RT "Thermolabile alanine racemase from a psychrotroph, Pseudomonas  
Biosci. Biotechnol. Biochem. 57:93-97(1993).  
CC -I- FUNCTION: CELL WALL FORMATION. IT IS THERMOLABILE AND SHOWS HIGH  
ACTIVITY AT LOW TEMPERATURES.  
CC -I- CATALYTIC ACTIVITY: L-ALANINE = D-ALANINE.

CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: ALONG WITH D-ALANINE-D-ALANINE LIGASE, IT MAKES UP THE  
 CC D-ALANINE BRANCH OF THE PEPTIDOGLYCAN BIOSYNTHETIC ROUTE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SIMILARITY: BELONGS TO THE ALANINE RACEMASE FAMILY.  
 DR PIR: PC1221; PC1221.  
 DR InterPro: IPR000821; Ala\_racemase.  
 DR PROSITE: PS00395; ALANINE RACEMASE, PARTIAL.  
 KM Isomerase; Pyridoxal phosphate; Cell wall; Peptidoglycan synthesis.  
 FT NON\_TER 25 25  
 FT SEQUENCE 25 AA: 2774 MW: 54D51109F62B1A43 CRC64.

Query Match 2.2%; Score 5; DB 1; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 IDLQA 102  
 Db 8 IDLQA 12

RESULT 3  
 ID ACON\_CANAL STANDARD; PRT; 28 AA.  
 AC P82611;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ACONITATE HYDROLASE, MITOCHONDRIAL (EC 4.2.1.3) (CITRATE HYDRO-LYASE)  
 DE (ACONITASE) (FRAGMENTS).  
 GN ACO1.  
 OS Candida albicans (Yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE, AND MASS SPECTROMETRY.  
 RC STRAIN=SC5314;  
 RX MEDLINE=20403418; PubMed=10949142;  
 RA Pardo M., Ward M., Pitarich A., Sanchez M., Nombela C., Blackstock W.,  
 RA Gil C.;  
 RT "Cross-species identification of novel *Candida albicans* immunogenic  
 RT proteins by combination of novel *Candida albicans* immunogenic  
 RT electrophoresis and mass spectrometry.";  
 RL Electrophoresis 21:2651-2659(2000).  
 CC -1- CATALYTIC ACTIVITY: CITRATE = CIS-ACONITATE + H(2)O.  
 CC -1- COFACTOR: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE (3FE-  
 CC 4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE CATALYTIC  
 CC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND ISOCITRATE (BY  
 CC SIMILARITY).  
 CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.  
 CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.  
 DR InterPro: IPR001030; Aconitase.  
 DR PROSITE: PS00450; ACONITASE\_1; PARTIAL.  
 DR PROSITE: PS01244; ACONITASE\_2; PARTIAL.  
 KM Lyase; Tricarboxylic acid cycle; Iron-sulfur; Mitochondrion; 4Fe-4S.  
 FT NON\_TER 1 1  
 FT METAL 7 7 IRON-SULFUR (4FE-4S) (BY SIMILARITY).  
 FT UNSURE 3 3 OR I.  
 FT UNSURE 4 4 OR I.  
 FT NON\_CONS 17 18  
 FT NON\_TER 28 28  
 FT SEQUENCE 28 AA: 3036 MW: C8D11B897D071291 CRC64;

Query Match 2.2%; Score 5; DB 1; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 PGSVQ 43

Db 22 PGSVQ 26

RESULT 4  
 ID HS98\_NEUCR STANDARD; PRT; 29 AA.  
 AC P31540;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-JUL-1993 (Rel. 26, Last annotation update)  
 DE HEAT SHOCK PROTEIN HSP98 (FRAGMENT).  
 OS Neurospora crassa.  
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 CC Sordariales; Sordariaceae; Neurospora.  
 OX NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=74A;  
 RX MEDLINE=93112646; PubMed=1472534;  
 RA Vassiliev A.O., Plesofsky-Vig N., Brambl R.;  
 RT "Isolation, partial amino acid sequence, and cellular distribution of  
 RT heat-shock protein hsp98 from *Neurospora crassa*.";  
 RL Biochim. Biophys. Acta 1156:1-6(1992).  
 CC -1- SUBCELLULAR LOCATION: MORE CONCENTRATED IN POLYRIBOSOMES THAN  
 CC IN MONORIBOSOMES, AND PREFERENTIALLY LOCALIZED IN THE LARGE  
 CC SUBUNIT.  
 CC -1- SIMILARITY: BELONGS TO THE CLP/A/CLPB FAMILY.  
 CC PIR: S28174; S28174.  
 DR InterPro: IPR001270; CLP\_AB.  
 DR PROSITE: PS00870; CLPAB\_1; 1.  
 DR PROSITE: PS00871; CLPAB\_2; PARTIAL.  
 KM Chaperone; Heat shock; ATP-binding.  
 FT NON\_TER 1 1  
 FT NON\_TER 29 29  
 FT SEQUENCE 29 AA: 3194 MW: 51A1BD58B81081B6 CRC64;

Query Match 2.2%; Score 5; DB 1; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 208 GATTL 212  
 Db 20 GATTL 24

RESULT 5  
 ID VAAL\_EQUAR STANDARD; PRT; 30 AA.  
 AC 004236;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT A, ISOFORM 1 (EC 3.6.1.34)  
 DE (FRAGMENT).  
 OS Equisetum arvense (Field horsetail) (Common horsetail).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Equisetophyta; Sphenopsida; Equisetales; Equisetaceae; Equisetum.  
 OX NCBI\_TaxID=3258;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=93138084; PubMed=8422915;  
 RA Starke T., Gogarten J.P.;  
 RT "A conserved intron in the V-ATPase A subunit genes of plants and  
 RT algae.";  
 RL FEBS Lett. 315:252-258(1993).  
 CC -1- FUNCTION: CATALYTIC SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF  
 CC VACUOLAR ATPASE. V-ATPASE VACUOLAR ATPASE IS RESPONSIBLE FOR  
 CC ACIDIFYING A VARIETY OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC  
 CC CELLS.  
 CC -1- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A  
 CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNITS A, B,

CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE  
 CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN).  
 CC -1- MISCELLANEOUS: TWO SEPARATE GENES ENCODE THE CATALYTIC 70 KDA  
 CC V-ATPASE SUBUNIT IN PSIIOTUM AND EUISETUM.  
 CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 DR EMBL: X556983; CAA40301.1; .  
 DR InterPro: IPR000194; ATPase\_alpha\_beta.  
 DR PROSITE: PS00152; ATPASE\_ALPHA\_BETA; PARTIAL.  
 KW ATP synthesis; Hydrogen ion transport; Hydrolyase; ATP-binding;  
 KW Multigene family.  
 FT NON\_TER 1 1  
 FT MOD\_RES 30 30  
 SQ SEQUENCE 30 AA: 3337 MW: 9627ED6206BD761F CRC64;

Query Match 2.2%; Score 5; DB 1; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 9 LTTLV 13  
 Db 25 LTTLV 29

RESULT 6  
 LITR\_PHYRO STANDARD; PRT: 9 AA.  
 AC P08946;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE RHODEI-LITORIN.  
 OS Phyllomedusa rohdei (Rhode's leaf frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
 OC Phyllomedusa.  
 OX NCBI\_TaxID=8394;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=85127560; PubMed=3838283;  
 RA Barra D., Erspamer G.F., Simmaco M., Bossa F., Melchiorri P.,  
 RA Erspamer V.;  
 RT "Rhodei-litorin: a new peptide from the skin of Phyllomedusa rohdei."  
 RL FEBS Lett. 182:53-56(1985).  
 CC -1- SIMILARITY: BELONGS TO THE BOMBESIN/NEURONEDIN B/RANATENSIN  
 CC FAMILY.  
 CC PIR: S07241.  
 DR InterPro: IPR000874; Bombesin.  
 DR Pfam: PF02044; Bombesin; 1.  
 DR PROSITE: PS00257; BOMBESIN; 1.  
 KW Bombesin family; Amidation.  
 FT MOD\_RES 1 1  
 FT MOD\_RES 9 9  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 9 AA: 1090 MW: 4ECCCE1861ADC377 CRC64;

Query Match 1.7%; Score 4; DB 1; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 218 LMAT 221  
 Db 2 LMAT 5

RESULT 7  
 TKL4\_LOCM1 STANDARD; PRT: 10 AA.  
 AC P30250;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE LOCUSTATACHYKININ IV (TK-IV).  
 OS Locusta migratoria (Migratory Locust).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;  
 OC Acridoidea; Acridoidea; Acrididae; Locusta.  
 OX NCBI\_TaxID=7004;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=91219696; PubMed=2132575;  
 RA Schoofs L., Holman G.M., Hayes T.K., Kochansky J.P., Nachman R.J.,  
 RA de Loof A.;  
 RT "Locustatachykinin III and IV: two additional insect neuropeptides  
 RT with homology to peptides of the vertebrate tachykinin family."  
 RL Regul. Pept. 31:199-212(1990).  
 CC -1- FUNCTION: MYOACTIVE PEPTIDE. STIMULATES THE CONTRACTION OF THE  
 CC OVIDUCT AND FOREGUT.  
 CC -1- SIMILARITY: SOME SIMILARITY TO TACHYKININS.  
 DR PIR: B60073; ECLQ4M.  
 KW Tachykinin; Neuropeptide; Amidation.  
 FT MOD\_RES 10 10  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 10 AA: 1040 MW: 9E52CD71E9C87735 CRC64;

Query Match 1.7%; Score 4; DB 1; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 186 GFHG 189  
 Db 5 GFHG 8

RESULT 8  
 TKSL\_AEDAE STANDARD; PRT: 10 AA.  
 AC P42634;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE SIALOKININ I.  
 OS Aedes aegypti (Yellowfever mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Aedes.  
 OX NCBI\_TaxID=7159;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=ROCKEFELLER; TISSUE=Salivary gland;  
 RX MEDLINE=94105119; PubMed=8278354;  
 RA Champagne D.E., Ribeiro J.M.C.;  
 RA "Sialokinin I and II: vasodilatory tachykinins from the yellow fever  
 RT mosquito Aedes aegypti."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:138-142(1994).  
 CC -1- FUNCTION: VASODILATORY PEPTIDE. MAY ACTIVATE MACROPHAGES AT THE  
 CC SITE OF FEEDING.  
 CC -1- SIMILARITY: SOME SIMILARITY TO TACHYKININS.  
 CC PIR: A49581; A49581.  
 DR InterPro: IPR002040; Tachykinin.  
 DR PROSITE: PS00267; TACHYKININ; 1.  
 KW Tachykinin; Neuropeptide; Amidation.  
 FT MOD\_RES 10 10  
 FT MOD\_RES 10 10  
 SQ SEQUENCE 10 AA: 1145 MW: 3DCFDE6B59C33AA8 CRC64;

Query Match 1.7%; Score 4; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 158 NTGD 161  
|||||

DB 1 NTGD 4

# RESULT 9

TRP8\_LEUMA STANDARD; PRT: 10 AA.  
ID TRP8\_LEUMA  
AC P81740;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE TACHYKININ-RELATED PEPTIDE 8 (LEMTIP 8).  
OS Leucophaea maderae (Madeira cockroach).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;  
OC Blaberidae; Blaberidae; Leucophaea.  
OX NCBI\_TaxID=6988;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE=Brain;  
RX MEDLINE=97269266; Pubmed=91144447;  
RA Muren J.E., Naessel D.R.;  
RT "Seven tachykinin-related peptides isolated from the brain of the Madeira cockroach: evidence for tissue-specific expression of isoforms.";  
RL Peptides 18:7-15(1997).  
RT FUNCTION: MYOACTIVE PEPTIDE. INCREASES THE AMPLITUDE AND FREQUENCY OF SPONTANEOUS CONTRACTIONS AND TONUS OF HINDGUT MUSCLE.  
CC -1- TISSUE SPECIFICITY: BRAIN.  
CC -1- MASS SPECTROMETRY: MW-1076.9; METHOD=MALDI.  
CC -1- SIMILARITY: SOME SIMILARITY TO TACHYKININS.  
CC Tachykinin; Neuropeptide; Amidation.  
KW MOD\_RES 10  
FT AMIDATION  
SQ SEQUENCE 10 AA; 1076 MW; 9E410371E9C87685 CRC64;

Query Match 1.7%; Score 4; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 186 GFHG 189  
|||||

DB 5 GFHG 8

# RESULT 10

FREL\_LITIN STANDARD; PRT: 12 AA.  
ID FREL\_LITIN  
AC P82021;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PRENATIN 1.  
OS Litoria infrafronata (giant tree frog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;  
OC Litoria.  
OX NCBI\_TaxID=61195;  
RN [1]  
RP SEQUENCE, AND MASS SPECTROMETRY.  
RC TISSUE=Skin secretion;  
RX MEDLINE=97368637; Pubmed=9225251;  
RA Raftery M.J., Waugh R.J., Bowie J.H., Wallace J.C., Tyler M.J.;  
RT "The structures of the frenatin peptides from the skin secretion of the giant tree frog Litoria infrafronata.";  
RL J. Pept. Sci. 2:117-124(1996).  
CC -1- FUNCTION: WIDE SPECTRUM ANTIMICROBIAL PEPTIDE.  
CC -1- TISSUE SPECIFICITY: SECRETED BY THE SKIN PAROTOID AND/OR ROSTRAL

CC GLANDS.  
CC -1- MASS SPECTROMETRY: MW-1140; METHOD=FAB.  
CC Antibiotic; Amphibian skin; Amidation.  
KW MOD\_RES 12  
FT AMIDATION  
SQ SEQUENCE 12 AA; 1141 MW; C62250BC365B72D CRC64;

Query Match 1.7%; Score 4; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 205 GILG 208  
|||||

DB 8 GILG 11

# RESULT 11

PORD\_METTM STANDARD; PRT: 12 AA.  
ID PORD\_METTM  
AC P80903;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PYRUVATE SYNTHASE SUBUNIT PORD (EC 1.2.7.1) (PYRUVATE OXIDOREDUCTASE DELTA CHAIN) (POR) (PYRUVIC-FERREDOXIN OXIDOREDUCTASE DELTA SUBUNIT) (FRAGMENT).  
CN PORD.  
OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).  
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
OC Methanothermobacter.  
OX NCBI\_TaxID=79929;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=97261844; Pubmed=9108258;  
RA Tensteegen A., Linder D., Thauer R.K., Hedderich R.;  
RT Structures and functions of four anaerobic 2-oxoacid oxidoreductases in Methanobacterium thermoautotrophicum.";  
RL Eur. J. Biochem. 244:862-868(1997).  
CC -1- FUNCTION: THE PH OPTIMUM IS PH 10.0 AND THE OPTIMAL TEMPERATURE IS 80 DEGREES CELSIUS.  
CC -1- CATALYTIC ACTIVITY: PYRUVATE + COA + OXIDIZED FERREDOXIN = ACETYL-COA + CO(2) + REDUCED FERREDOXIN.  
CC -1- COFACTOR: BINDS TWO 4FE-4S CLUSTERS (BY SIMILARITY).  
CC -1- SUBUNIT: HETEROTETRAMER OF ONE ALPHA, ONE BETA, ONE DELTA AND ONE GAMMA CHAIN.  
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL TYPE FERREDOXIN FAMILY.  
CC Oxidoreductase; Electron transport; Iron-sulfur Repeat; 4fe-4s.  
KW NON\_TER 12  
FT 12  
SQ SEQUENCE 12 AA; 1241 MW; 2D54065D1BD1ADD8 CRC64;

Query Match 1.7%; Score 4; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 1.6e+03;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 207 LGAT 210  
|||||

DB 4 LGAT 7

# RESULT 12

V25K\_MSSV STANDARD; PRT: 12 AA.  
ID V25K\_MSSV  
AC P82004;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE 25 KDA STRUCTURAL POLYPROTEIN (FRAGMENT).  
OS White spot syndrome virus (WSSV).  
OC Viruses; Unassigned viruses.  
OX NCBI\_TaxID=92652;  
RN [1]  
RP SEQUENCE.

RC STRAIN-South Carolina;  
 RX MEDLINE=20214217; PubMed=10752552;  
 RA Wang Q., Poulos B.T., Lightner D.V.;  
 RT "Protein analysis of geographic isolates of shrimp white spot syndrome  
 RT virus.";  
 RL Arch. Virol. 145:263-274(2000).  
 CC -1- FUNCTION: STRUCTURAL COMPONENT OF THE VIRION.  
 FT NON\_TER 12 12  
 SQ SEQUENCE 12 AA; 1283 MW; C5409AD9ECB731A9 CRC64;

Query Match  
 Best Local Similarity 1.7%; Score 4; DB 1; Length 12;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 211 TSLV 214  
 DB 6 TSLV 9

RESULT 13  
 CHEP\_PARID STANDARD; PRT; 13 AA.  
 AC P42718;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DE 01-NOV-1995 (Rel. 32, Last annotation update)  
 OS CHEMOTACTIC PEPTIDE.  
 CC Parapolybia indica.  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;  
 CC Vespoidea; Vespidae; Polistinae; Parapolybia.  
 OX NCBI\_Taxid=31921;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Venom;  
 RA Toki T., Yasuhara T., Nakajima T.;  
 RT "Isolation and sequential analysis of peptides on the venom sac of  
 RT Parapolybia indica.";  
 RL Eisei Dobutsu 39:105-111(1968).  
 KW Chemotaxis; Amidation.  
 FT MOD\_RES 13 13  
 SQ SEQUENCE 13 AA; 1298 MW; 5C950CE8E39D5873 CRC64;

Query Match  
 Best Local Similarity 1.7%; Score 4; DB 1; Length 13;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 ILGL 75  
 DB 1 ILGL 4

RESULT 14  
 KLOM\_LUMTE STANDARD; PRT; 15 AA.  
 AC P11918;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE LUMBRICINE KINASE (EC 2.7.3.5) (UK) (FRAGMENT).  
 OS Lumbricus terrestris (Common earthworm).  
 CC Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
 CC Lumbricina; Lumbricidae; Lumbricus.  
 OX NCBI\_Taxid=6398;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=7206544; PubMed=5128744;  
 RA der Terrossian E., Desvages G., Pradel L.A., Kassab R., van Thoi N.;  
 RT "Comparative structural studies of the active site of ATP: guanidine  
 RT phosphotransferases. The essential cysteine tryptic peptide of  
 RT lumbricine kinase from Lumbricus terrestris muscle.";

RL Eur. J. Biochem. 22:585-592(1971).  
 CC -1- CATALYTIC ACTIVITY: ATP + LUMBRICINE = ADP + N-PHOSPHOLUMBRICINE.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- SIMILARITY: BELONGS TO THE ATP:GUANIDO PHOSPHOTRANSFERASE FAMILY.  
 DR PIR: A08416; A08416.  
 DR InterPro: IPR000749; ATP\_gua\_Ptrans.  
 DR PROSITE: PS00112; GUANIDO\_KINASE; 1.  
 KW Kinase; Transferase.  
 FT NON\_TER 1 1  
 FT ACT\_SITE 6 6  
 FT NON\_TER 15 15  
 SQ SEQUENCE 15 AA; 1565 MW; 2AA5FEF6140E90C4 CRC64;

Query Match  
 Best Local Similarity 1.7%; Score 4; DB 1; Length 15;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 CPGS 41  
 DB 6 CPGS 9

RESULT 15  
 PGKH\_PHYPA STANDARD; PRT; 15 AA.  
 AC P80659;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PHOSPHOGLYCERATE KINASE, CHLOROPLAST  
 OS Physcomitrella patens (Moss).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 CC Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.  
 OX NCBI\_Taxid=3218;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Protonema;  
 RX MEDLINE=97275459; PubMed=9129336;  
 RA Kasten B., Buck F., Nuske J., Reski R.;  
 RT "Cytokinin affects nuclear- and plastome-encoded energy-converting  
 RT plastid enzymes.";  
 RL Planta 201:261-272(1997).  
 CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +  
 CC 3-PHOSPHO-D-GLYCERYL PHOSPHATE.  
 CC -1- PATHWAY: CALVIN CYCLE.  
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.  
 DR InterPro: IPR001576; PGK.  
 DR PROSITE: PS00111; PGKINASE\_KINASE; PARTIAL.  
 KW transferase; kinase; multigene family; Calvin cycle; Chloroplast.  
 FT NON\_TER 15 15  
 SQ SEQUENCE 15 AA; 1531 MW; 56A5EC1F677EBC6 CRC64;

Query Match  
 Best Local Similarity 1.7%; Score 4; DB 1; Length 15;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 46 SKVA 49  
 DB 7 SKVA 10

Search completed: March 28, 2002, 09:25:36  
 Job time: 534 sec

Thu Mar 28 09:21:31 2002

us-09-726-348-2.rsp

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 08:58:40 ; Search time 105.47 Seconds

(Without alignments)  
124.310 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_1\_177

Perfect score: 971

Sequence: 1 MAPHGPGSLTLTPWMAALL.....NTGPEKPCENGSCVPPGPG 177

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: /SIDS1/gcgdata/geneseq/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	971	100.0	229	18	AAW27087
2	971	100.0	229	18	AAW13944
3	971	100.0	229	21	AAW33419
4	971	100.0	229	21	AAW88570
5	971	100.0	229	21	AAW04295
6	971	100.0	229	22	AAW20112
7	971	100.0	229	22	AAW88595
8	971	100.0	229	22	AAW50951
9	971	100.0	229	22	AAW48068
10	959	98.8	229	20	AAW05282
11	790	81.4	147	21	AAW00157

12	163	16.8	52	22	AAW04296	Transforming growth
13	99	10.2	298	21	AAW07469	A human leucine-ri
14	94.5	9.7	281	22	AAW02198	Human PRO1341 poly
15	84.5	8.7	385	16	AAW60176	Chimeric protein (
16	82	8.4	670	16	AAW67759	Human fetal brain
17	82	8.4	769	16	AAW75352	Human fetal brain
18	81	8.3	432	20	AAW88236	Mouse prothrombina
19	80.5	8.3	1059	22	AAW00448	Maize disease resl
20	80	8.2	802	22	AAW41332	Human polypeptide
21	79.5	8.2	321	19	AAW53243	Mus musculus vascu
22	79.5	8.2	337	20	AAW08286	Human growth facto
23	79.5	8.2	358	19	AAW53242	Mus musculus vascu
24	79.5	8.2	358	19	AAW44295	Mouse vasculat end
25	78.5	8.1	806	17	AAW09254	Tomato pathogen re
26	78.5	8.1	806	17	AAW85301	L. blmstus plasmi
27	77	7.9	296	20	AAW27073	Human polypeptide
28	77	7.9	751	22	AAW39546	Rat slit homologue
29	76.5	7.9	1529	21	AAW76117	Rabbit LDL recepto
30	76.5	7.9	1529	21	AAW60556	Human protein sequ
31	76.5	7.9	2213	18	AAW63556	Human scph3.10 (SA
32	76	7.8	445	22	AAW93714	Rat vascular endot
33	76	7.8	571	20	AAW44296	Murine c-Fos induc
34	75.5	7.8	326	19	AAW4296	Human protein. Xen
35	75.5	7.8	358	18	AAW44992	Human EMRI seven t
36	75.5	7.8	1530	21	AAW28151	Rice COT1 protein
37	75	7.7	692	21	AAW25522	Pinus radiata cell
38	75	7.7	880	21	AAW01249	Human EMRI seven t
39	75	7.7	880	21	AAW71869	Rice COT1 protein
40	74.5	7.7	597	21	AAW3457	Pinus radiata cell
41	74.5	7.7	804	21	AAW35515	Amino acid sequenc
42	72.5	7.5	217	19	AAW37863	Human T1 receptor
43	72.5	7.5	217	19	AAW48334	Human enterokinase
44	72.5	7.5	1019	22	AAW06940	Chick Serrate. Ga
45	72.5	7.5	1193	17	AAW05835	

## ALIGNMENTS

RESULT 1	AAW27087	standard; Protein: 229 AA.
ID	AAW27087	
AC	AAW27087	
XX		
DT	28-JAN-1998	(first entry)
XX		
DE	Human transforming growth factor alpha HIII.	
XX		
KW	human transforming growth factor; TGF-alpha-HIII; angiogenesis;	
KW	embryogenesis; ocular disorder; kidney disorder; liver disorder;	
KW	neural disorder; alopecia; inflammation.	
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..25
FT		/label= signal-peptide
FT		/note= "putative"
FT	Protein	26..229
FT		/label= mature_protein
FT	Region	1..177
FT		/label= soluble_portion_of_protein
FT	Region	126..177
FT		/label= active_site
FT	Region	178..204
FT		/note= "transmembrane_portion
PN		/note= "putative"
XX	MO9725349-AA1.	
XX	17-JUL-1997.	

PF 04-JAN-1996: 96WO-US00149.  
 XX  
 PR 04-JAN-1996: 96WO-US00149.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Wei Y;  
 XX  
 DR WPI: 1997-372817/34.  
 DR N-PSDB: AAT85082.  
 XX  
 PT New human transforming growth factor-alpha homologue - used for  
 PT developing products for treating e.g. neurological disorders, kidney  
 PT and liver disorders, tumours, wounds, hair loss or skin disorders  
 XX  
 PS Claim 15; Page 47; 63pp; English.

CC This protein has been putatively identified as a human transforming  
 CC growth factor (TGF) alpha analogue, TGF-alpha-HII. The protein can  
 CC stimulate angiogenesis, embryogenesis, cell differentiation and function.  
 CC It can be used for therapeutic purposes for restoration or enhancement of  
 CC neurological functions diminished as a result of trauma or other damaging  
 CC pathological functions such as AIDS dementia and senile dementia, to treat ocular  
 CC disorders, e.g. corneal inflammation, to destroy target cells, to treat  
 CC tumours, kidney or liver disorders or to treat wounds, burns or ulcers.  
 CC The polypeptide can also be used in the modulation of angiogenesis, bone  
 CC resorption, immune response, and synaptic and neuronal effector  
 CC functions, or the arachidonic acid cascade. It can also be used in  
 CC applications related to terminal differentiation e.g. in  
 CC hyperproliferative disorders such as inflammation or psoriasis and for  
 CC alopecia, hair loss or other skin conditions which affect hair follicular  
 CC development. Antagonists to TGF-alpha-HII can be used for treating  
 CC tumours or skin disorders such as psoriasis. The products can also  
 CC be used for diagnosis and detection of the above disorders.

XX Sequence 229 AA:

Query Match 100.0%; Score 971; DB 18; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-93;  
 Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPHGSGITLVPMNAALLALGVERALALPEICTOCPSGVONLSKVAFCYKTRRELM 60  
 DB 1 maphgyslttlvpmnaallalgyeralalpelictcpsgvnlskvaifcxtrelml 60  
 OY 61 HARCCLNOKGTLIGLDLONCSLEDPGPNFHQAHTTVIIDLANPLKGLANTPRGFTQLO 120  
 DB 61 harccclngkgtllgldlqncsleddpgpnfhqahttviiidlanplkglantprgftq 120  
 OY 121 TLLPQHVCNCPGGINAMNTTSTYDNOICOGKNCNNTGDEMCPENGSCVPDPG 177  
 DB 121 tllpqhvcncpgginamnttstydnoicogkncnntgdepemcpeengscvpdp 177

RESULT 2  
 AAY13944 standard; Protein: 229 AA.

XX AAY13944:  
 XX  
 DT 14-JUL-1999 (first entry)

XX Human transmembrane protein, HP10435.  
 DE  
 XX  
 KW Transmembrane protein; human; cell membrane; proliferation; diagnosis;  
 KW cell differentiation; carcinostatic agent; probe; gene therapy;  
 KW signal transduction; apoptosis; inhibitor;  
 KW phosphatidylethanolamine N-methyltransferase.

OS Homo sapiens.  
 XX  
 PN W09918203-A2.

XX  
 PD 15-APR-1999.  
 XX  
 PF 05-OCT-1998: 98WO-JP04475.  
 XX  
 PR 08-OCT-1997: 97JP-0276271.  
 XX  
 PA (PROT-) PROTEGENE INC.  
 PA (SAGA) SAGAMI CHEM RES CENT.  
 XX  
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
 XX  
 DR WPI: 1999-277268/23.  
 DR N-PSDB: AAX36812, AAX36813.  
 XX  
 PT Human transmembrane proteins and nucleotide sequences  
 PT  
 PS Claim 1; Page 90-91; 139pp; English.

CC This sequence is a human transmembrane protein of the invention.  
 CC All of the proteins exist in the cell membrane, so are considered to be  
 CC proteins controlling the proliferation and differentiation of the cells.  
 CC They may be useful as carcinostatic agents or as antigens for preparing  
 CC antibodies against the proteins. The cDNAs can be used as probes for  
 CC gene diagnosis and gene sources for gene therapy, as well as for  
 CC large-scale expression of the proteins. The HP01498 (see AAY13939)  
 CC protein may be associated with signal transduction associated with  
 CC apoptosis, and therefore useful in inhibition of apoptosis. The HP01962  
 CC (see AAY13943) protein can be used to treat diseases associated with  
 CC phosphatidylethanolamine N-methyltransferase. The proteins are  
 CC identified by the presence of a hydrophobic transmembrane region,  
 CC knowledge of the protein function is not required, as in e.g. methods of  
 CC expression cloning.

XX Sequence 229 AA:

Query Match 100.0%; Score 971; DB 20; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-93;  
 Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPHGSGITLVPMNAALLALGVERALALPEICTOCPSGVONLSKVAFCYKTRRELM 60  
 DB 1 maphgyslttlvpmnaallalgyeralalpelictcpsgvnlskvaifcxtrelml 60  
 OY 61 HARCCLNOKGTLIGLDLONCSLEDPGPNFHQAHTTVIIDLANPLKGLANTPRGFTQLO 120  
 DB 61 harccclngkgtllgldlqncsleddpgpnfhqahttviiidlanplkglantprgftq 120  
 OY 121 TLLPQHVCNCPGGINAMNTTSTYDNOICOGKNCNNTGDEMCPENGSCVPDPG 177  
 DB 121 tllpqhvcncpgginamnttstydnoicogkncnntgdepemcpeengscvpdp 177

RESULT 3  
 AAB33419 standard; Protein: 229 AA.

XX AAB33419:  
 XX  
 DT 29-JAN-2001 (first entry)

XX Human PRO240 protein UNO214 SEQ ID NO:26.  
 DE  
 XX  
 KW Human; immune related disease; diagnosis; antiinflammatory; cardiant;  
 KW dermatological; antiallergic; antiinflammatory; immunosuppressive;  
 KW haemostatic; antithyroid; antidiabetic; neurotropic; neuroprotective;  
 KW antianemic; hepatotropic; virucide; antiproliferative; antiallergic;  
 KW antianemic; systemic lupus erythematosus; rheumatoid arthritis;  
 KW osteoarthritis; spondyloarthritis; systemic sclerosis; sarcoidosis;  
 KW idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;  
 KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;  
 KW autoimmune thrombocytopenia; immune-mediated renal disease;



KW demyelinating disease; hepatobiliary disease; Whipple's disease;  
 KW inflammatory bowel disease; gluten-sensitive enteropathy;  
 KW autoimmune disease; immune-mediated skin disease; allergic disease;  
 KW immunological disease; transplantation associated disease;  
 KW graft rejection; graft-versus-host-disease.  
 OS Homo sapiens.  
 XX WO200053758-A2.  
 PN 14-SEP-2000.  
 PD 02-MAR-2000; 2000WO-US05841.  
 PF 08-MAR-1999; 99WO-US05028.  
 PR 10-MAR-1999; 99US-0123618.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 23-MAR-1999; 99US-0125775.  
 PR 12-APR-1999; 99US-0128849.  
 PR 20-APR-1999; 99WO-US08615.  
 PR 28-APR-1999; 99US-0131445.  
 PR 04-MAY-1999; 99US-0132371.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 23-JUN-1999; 99US-01441037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 28-JUL-1999; 99US-0146222.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21547.  
 PR 05-OCT-1999; 99WO-US23089.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 29-NOV-1999; 99WO-US28214.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 30-NOV-1999; 99WO-US28409.  
 PR 01-DEC-1999; 99WO-US28301.  
 PR 02-DEC-1999; 99WO-US28634.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28564.  
 PR 16-DEC-1999; 99WO-US28565.  
 PR 20-DEC-1999; 99WO-US30095.  
 PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 11-FEB-2000; 2000WO-US00376.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 22-FEB-2000; 2000WO-US04342.  
 XX (GETH ) GENENTECH INC.  
 XX Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W,  
 PI Kabakoff RC, Lu Y, Pan J, Pennica D, Shelton DL, Smith V;  
 PI Stewart TA, Tumes D, Watanabe CK, Wood WI, Yan M;  
 DR WPI: 2000-572271/53.  
 DR N-PSDB: AAC58584.  
 XX Sixty four PRO polypeptides, useful in the diagnosis and treatment of  
 PT immune related disorders, e.g. systemic lupus erythematosus, rheumatoid  
 PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -  
 XX Claim 33; Fig 12; 309pp; English.  
 XX The present invention describes sixty four human PRO proteins which can  
 CC be used in the treatment of immune related diseases. The human PRO  
 CC proteins, anti-PRO antibodies, agonists and antagonists are useful for  
 CC treating and diagnosing immune related disorders. The disorders are

CC selected from systemic lupus erythematosus, rheumatoid arthritis,  
 CC osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,  
 CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's  
 CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic  
 CC anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus,  
 CC immune-mediated renal disease, demyelinating diseases of the central  
 CC and peripheral nervous systems, hepatobiliary diseases, inflammatory  
 CC bowel disease, gluten-sensitive enteropathy and Whipple's disease,  
 CC autoimmune or immune-mediated skin diseases, allergic diseases,  
 CC immunological diseases of the lung, and transplantation associated  
 CC diseases including graft rejection and graft-versus-host-disease.  
 CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used  
 CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and  
 CC AAC33414 to AAC33477 represent human PRO polynucleotide and protein  
 CC sequences given in the exemplification of the present invention.  
 XX Sequence 229 AA;  
 SQ  
 Query Match 100.0%; Score 971; DB 21; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 5,1e-93;  
 Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAPHGPSLTTLVPMARALALGVERALALPEICQCRCGSVQNSLKAFFYKTTREML 60  
 Db 1 maphgpslttlvpmaraallalqveratalpeictcpgsvqnskvaifyektrelml 60  
 QY 61 HARCLLNQKGTILGLDQNSLEDPGPFHOATTVYIIDLQANPLKGDLANFRGFQLO 120  
 Db 61 harccllnqkgtilglldqnsledpgpntfgahtrvliidqanplkgdlanfriftq 120  
 QY 121 TLILPHVNCPCGINMNTTTSYIDNQCQGNLCNNTGDEPCPCPCVDPGPG 177  
 Db 121 tllipqhvncpgginmntttsyidnqicqgknlcnntgdcpcpcpcvdpdpg 177  
 RESULT 4  
 AAY88570  
 ID AAY88570 standard; Protein: 229 AA.  
 XX  
 AC AAY88570;  
 DT 09-AUG-2000 (first entry)  
 DT  
 DE Human PRO240 amino acid sequence.  
 KW Antibody: PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;  
 KW PRO317; tumour growth inhibitor; cancer; diagnosis; treatment; human;  
 KW cell growth proliferation; serrate precursor; C-serrate-1; ADEPT;  
 KW antibody dependent enzyme mediated prodng therapy; chromosome 2.  
 OS Homo sapiens.  
 XX  
 PN WO200015666-A2.  
 XX 23-MAR-2000.  
 PD 08-SEP-1999; 99WO-US20594.  
 PF 10-SEP-1998; 98US-0099803.  
 PR 10-SEP-1998; 98WO-US18824.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 XX Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI, Botstein D;  
 DR WPI: 2000-271386/23.  
 DR N-PSDB: AAA30036.  
 XX New isolated antibodies which bind to specific polypeptides used for  
 PT diagnosis and treatment of neoplastic cell growth and proliferation -  
 PT Example 4; Fig 8; 200pp; English.

PD

PRO240; UNQ214; human; immune disease; autoimmune disease;

CC	cells into a tissue of a mammal, stimulating or enhancing an immune response, or increasing the proliferation of T-lymphocytes in a
CC	mammal in response to an antigen. Claimed compositions comprising
CC	a PRO polypeptide or its antagonist have the opposite effect. A
CC	claimed method for treating an immune related disorder, such as a T
CC	cell disorder, involves administering a PRO polypeptide, an agonist
CC	antibody or an antagonist antibody. The disorder is selected from,
CC	systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,
CC	idiopathic chronic arthritis, spondyloarthropathy, systemic sclerosis,
CC	vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune
CC	thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated
CC	renal disease, demyelinated diseases (such as multiple sclerosis),
CC	autoimmune chronic active hepatitis, primary biliary cirrhosis,
CC	granulomatous hepatitis, sclerosing cholangitis, inflammatory bowel
CC	disease (ulcerative colitis and Crohn's disease), gluten-sensitive
CC	enteropathy, Whipple's disease, (auto)immune-mediated skin diseases
CC	(such as bullous skin disease, erythema multiforme and psoriasis),
CC	allergic diseases (such as asthma, allergic rhinitis and psoriasis),
CC	dermatitis, food hypersensitivity and urticaria), idiopathic
CC	diseases of the lung and transplantation associated diseases (such
CC	as graft rejection and graft-versus-host disease) (all claimed).
CC	Claimed methods of diagnosing these disorders comprise detecting
CC	the level of expression of the PRO gene. Also claimed are a method
CC	of identifying a compound capable of inhibiting the expression or
CC	activity of the PRO polypeptide, vectors, host cells, antibodies,
CC	and a method of stimulating an immune response in a mammal using
CC	PRO240.
XX	
SQ	Sequence 229 AA;
Query Match	100.0%; Score 971; DB 22; Length 229;
Best Local Similarity	100.0%; Pred. No. 5,1e-93;
Matches 177; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	I MAPHGPGSLTTLTPMAAALLALGVERALAPDETCTOCPSGVOMLSKVAFCYCKTRRLMI 60
Dd	1 maphgpgslctltpwaaalllalgveralaalpeltcpgsvqnlksvalykxtrelml 60
QY	61 HARCCLNQKRTIIIGDLQWCSLEDPENFHOAHVTIIDLQANPLKGDLANTFRGFTOIQ 120
Dd	61 harccclnqkrtlllgdlqnscledcpnfhgahtvtviiidqlanplkgdiantfrgttqg 120
QY	121 TLIDPQHVNCPGSINAMNTITSYIDNOICCGCKNLCAINTGPPEMCPMGSCVVDGGC 177
Dd	121 tlidpqhvncpgsinamntitsyidnglcgqknlcmtgdpmcpcngscvypdpg 177
RESULT	7
AAB68595	AAB68595 standard; Protein: 229 AA.
XX	
AC	AAB68595;
XX	
DT	27-APR-2001 (first entry)
XX	
DE	PRO240.
OS	Homo sapiens.
KM	Cytostatic; PRO protein; tumour; cancer.
XX	
PN	WO200105836-A1.
PD	25-JAN-2001.
PF	20-DEC-1999; 99WO-US30999.
PR	20-JUL-1999; 99US-0144758.
R	26-JUL-1999; 99US-0145698.
RR	08-SEP-1999; 99WO-US20594.
RR	13-SEP-1999; 99WO-US20944.

PR 15-SEP-1999; 99MO-US21090.  
 PR 05-OCT-1999; 99MO-US23089.  
 PR 29-NOV-1999; 99MO-US28214.  
 PR 30-NOV-1999; 99MO-US28313.  
 PR 02-DEC-1999; 99MO-US28564.  
 XX (GETH ) GENENTECH INC.  
 XX Botstein D, Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI;  
 XX WPI: 2001-091968/10.  
 DR N-PSDB; AAF60356.  
 DR WPI: 2001-091968/10.  
 PR New antibody that binds to a PRO polypeptide, e.g. PRO187 and PRO533,  
 PT useful for diagnosing and treating cancers -  
 PS Claim 61; Fig 8; 196pp; English.  
 XX The present invention relates to PRO proteins and coding sequences. The  
 CC present sequence is one such PRO protein. It was found that the PRO genes  
 CC are amplified in the genome of tumour cells. The gene amplification is  
 CC expected to be associated with the overexpression of the gene product and  
 CC contributes to tumorigenesis. Therefore, antagonists of PRO proteins are  
 CC useful for the treatment of benign or malignant tumours, leukemias,  
 CC lymphoid malignancies and other disorders such as neuronal, glial,  
 CC astrocytal, hypothalamic, glandular, epithelial, inflammatory and  
 CC immunologic disorders.  
 XX Sequence 229 AA;  
 SQ

Query Match 100.0%; Score 971; DB 22; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 5, 1e-93;  
 Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAPHGSSTTLVPMMAALLALGYERALALPEICTGCGSVQNSKVAFCYCKTTRRLML 60  
 DB 1 maphgsgsttlvpmmaallalgyeralalpeictgcpvsgvskvafcyckttreml 60  
 QY 61 HARCCNLQKGTILGLDQNCSEDEPGPNFHOAHNTVIIDLANPLKGLDANTFRGFTQLQ 120  
 DB 61 harccnlqgkgtilglidqncsledepgpnfhqahntvtiiddlanplkglantfrgftqlq 120  
 QY 121 TLILPOHVNCPGGINAMNTTSTYIDNQICQGNKLCNNTGDEPMCPENGSCVPDGP 177  
 DB 121 tlilpqhvnpcpgginamnttstydinqicgqgnklnclntgdepemcpengscvpdpg 177

RESULT 8  
 AAB50951 standard; Protein: 229 AA.  
 ID AAB50951  
 XX AAB50951:  
 AC 21-MAR-2001 (first entry)  
 XX Human PRO240 protein.  
 DE Human PRO240 protein.  
 XX Human; PRO; cytosolic; nootropic; neuroprotective; respiratory general;  
 KW antiinflammatory; antiangiogenic; immunosuppressive; immunostimulant;  
 KW PRO agonist; cancer; inflammatory disorder; immunological disorder.  
 XX Homo sapiens.  
 OS  
 XX WO200073348-A2.  
 PN  
 XX 07-DEC-2000.  
 PD  
 XX 30-MAY-2000; 2000MO-US14941.  
 PF  
 XX 02-JUN-1999; 99MO-US12252.  
 PR 22-JUN-1999; 99US-0140650.  
 PR 23-JUN-1999; 99US-0141037.

PR 20-JUL-1999; 99US-0144758.  
 PR 01-SEP-1999; 99MO-US20111.  
 PR 08-SEP-1999; 99MO-US20594.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99MO-US28313.  
 PR 01-DEC-1999; 99MO-US28634.  
 PR 02-DEC-1999; 99MO-US28551.  
 PR 16-DEC-1999; 99MO-US30095.  
 PR 20-DEC-1999; 99MO-US30999.  
 PR 06-JAN-2000; 2000MO-US00376.  
 PR 11-FEB-2000; 2000MO-US03565.  
 PR 18-FEB-2000; 2000MO-US04342.  
 PR 18-FEB-2000; 2000MO-US05841.  
 PR 02-MAR-2000; 2000US-0187202.  
 PR 03-MAR-2000; 2000MO-US06319.  
 PR 10-MAR-2000; 2000MO-US06884.  
 PR 15-MAR-2000; 2000MO-US08439.  
 PR 30-MAR-2000; 2000MO-US13705.  
 PR 17-MAY-2000; 2000MO-US13705.  
 XX (GETH ) GENENTECH INC.  
 XX Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W, Kabakoff RC;  
 XX Shelton DL, Smith V, Watanabe CK, Wood WI;  
 XX WPI: 2001-016509/02.  
 DR N-PSDB; AAC91553.  
 XX Twenty eight nucleic acids encoding PRO polypeptides which are useful  
 PT for treating various tumors, e.g. breast cancer, and other  
 PT inflammatory, angiogenic and immunological disorders -  
 PS Claim 31; Fig 2; 188pp; English.  
 XX The present sequence is one of twenty eight novel PRO polypeptides. The  
 CC PRO polypeptides and their agonists, including antibodies, peptides, and  
 CC small molecule agonists, may be used to treat various tumors, e.g.,  
 CC cancers such as breast cancer, ovarian cancer, renal cancer, colorectal  
 CC cancer, uterine cancer, prostate cancer, lung cancer, bladder cancer,  
 CC central nervous system cancer, melanoma or leukemia. They are also  
 CC useful for treating other disorders such as neuronal, glial, astrocytal,  
 CC hypothalamic and other glandular, macrophagal, epithelial, stromal and  
 CC blastocoele disorders, and inflammatory, angiogenic and immunological  
 CC disorders.  
 XX Sequence 229 AA;  
 SQ

Query Match 100.0%; Score 971; DB 22; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 5, 1e-93;  
 Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAPHGSSTTLVPMMAALLALGYERALALPEICTGCGSVQNSKVAFCYCKTTRRLML 60  
 DB 1 maphgsgsttlvpmmaallalgyeralalpeictgcpvsgvskvafcyckttreml 60  
 QY 61 HARCCNLQKGTILGLDQNCSEDEPGPNFHOAHNTVIIDLANPLKGLDANTFRGFTQLQ 120  
 DB 61 harccnlqgkgtilglidqncsledepgpnfhqahntvtiiddlanplkglantfrgftqlq 120  
 QY 121 TLILPOHVNCPGGINAMNTTSTYIDNQICQGNKLCNNTGDEPMCPENGSCVPDGP 177  
 DB 121 tlilpqhvnpcpgginamnttstydinqicgqgnklnclntgdepemcpengscvpdpg 177

RESULT 9  
 AAB48068 standard; protein: 342 AA.  
 ID AAB48068  
 XX AAB48068:  
 AC 19-MAR-2001 (first entry)  
 XX  
 XX  
 XX  
 XX

DE Human extracellular signaling molecule (EXCS) (ID 2207183CD1).

XX Extracellular signaling molecule; EXCS; anti-inflammatory; human;

XX immunosuppressive; cytostatic; neuroprotective; gastrointestinal;

KW virus; anti-bacterial; anti-HIV; human immunodeficiency virus;

KW antileukemic; cerebroprotective; neurotropic; antitumor; antifungal;

KW anticonvulsant; tranquilizer; neuroleptic; vasotropic; gynecological;

KW keratolytic; protozoacide; gene therapy.

OS Homo sapiens.

XX

PN WO200070049-A2.

PD

XX 23-NOV-2000.

XX

PF 19-MAY-2000; 2000WO-US13975.

XX

PR 19-MAY-1999; 99US-0134949.

PR 15-JUL-1999; 99US-0144270.

PR 30-JUL-1999; 99US-0146700.

PR 04-OCT-1999; 99US-0157508.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Tang YT, Yue H, Lal P, Burford N, Bandman O, Baughn MR;

PI Azimzai Y, Lu DAM, Patterson C;

DR WPI: 2001-025021/03.

DR N-PSDB; AAC84304.

XX

PT New human extracellular signaling nucleic acids and polypeptides useful

PT for diagnosing, treating and preventing infections and

PT gastrointestinal, neurological, reproductive, and

PT autoimmune/inflammatory disorders -

XX

PS Claim 1; Page 89-90; 114pp: English.

XX

XX The invention provides human extracellular signaling molecules (EXCS)

CC and polynucleotides which identify and encode EXCS. EXCS can be

CC expressed by standard recombinant methodology. The amino acid and nucleic

CC acid sequences of EXCS are useful for diagnosing, treating and

CC preventing infections and gastrointestinal (peptic ulcer, dysphagia,

CC pancreatitis), neurological (e.g. epilepsy, ischemic cerebrovascular

CC disease, stroke), reproductive (infertility, ovulatory defects,

CC endometriosis), autoimmune/inflammatory (actinic keratosis, acquired

CC immunodeficiency syndrome (AIDS), Addison's disease), and cell

CC proliferative disorders including cancers (of the breast, adrenal gland,

CC bone). They may also be used to treat fatal familial insomnia,

CC nutritional and metabolic diseases of the nervous system, myopathies,

CC mental disorders (anxiety, schizophrenia, mood), as well as infections

CC caused by parasites (malaria, leishmania, trypanosoma), viral

CC (adenovirus, coronavirus, flavivirus), bacterial (e.g. pneumococcus,

CC staphylococcus, bacillus), and fungal (aspergillus, blastomycosis,

CC dermatophytes) agents. The nucleic acids, polypeptides, antagonists,

CC agonists, pharmaceutical compositions, and antibodies may also be

CC used for treating or preventing disorders associated with increased or

CC decreased expression or activity of EXCS. EXCS polynucleotides may also

CC be used to detect and quantify gene expression in biopsied tissues in

CC which expression of EXCS may be correlated with the disease, to determine

CC presence or excess expression of EXCS, to monitor regulation of EXCS

CC levels during therapeutic intervention, to detect the presence of

CC associated disorders, as targets in microarray, to generate hybridization

CC probes, and to detect differences in gene sequences among normal, carrier

CC or affected individuals. Antibodies may also be used in diagnosing

CC disorders, in monitoring patients being treated with EXCS agonists,

CC antagonists or inhibitors. Sequences AAB48057-B48082 represent the EXCS

CC of the invention.

XX

Sequence 342 AA:

Query Match 100.0%; Score 971; DB 22; Length 342;

Best Local Similarity 100.0%; Pred. No. 8,6e-93;

Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPHGSLTITVPMMAAILLALGVERALALPEICTGCGSVQNSKVAFCRTTRELML 60

DB 114 maphgsltltvpmmaailalgyeralalpeictcgpsvqnskvaifcrtrelml 173

OY 61 HARCCLNOKTTLGLDILONCSLEDPGNFQAHTVVIDLQANPKGLAMTFRGFTQLQ 120

DB 174 harcclngkgtllglldlqncslcdpgnfhgntltvldlqanplkgdiantfrgftqlq 233

OY 121 TLILPQHVNCPGGINMWNITTSYIDNQCCKNLGNNTGDPKCPENGSCVPDGP 177

DB 234 tlilpqhvnpcpgginawmltsyidnqcgqknlcnntgdpemcengscvpdp 290

RESULT 10

AAV05282

ID AAV05282 standard; Protein; 229 AA.

XX

AC AAV05282;

XX

DT 22-JUN-1999 (first entry)

XX

DE EGF-like homologue PRO240.

XX

KW Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;

KW EBAF-2; inhibitor; tumour growth; cancer; EGF-like homologue;

KW EBAF-8 homologue.

XX

OS Homo sapiens.

XX

PN WO9914327-A2.

XX

PD 25-MAR-1999.

XX

PF 10-SEP-1998; 98WO-US18824.

XX

PR 25-NOV-1997; 97US-0066840.

PR 17-SEP-1997; 97US-0059114.

PR 17-SEP-1997; 97US-0059117.

PR 18-SEP-1997; 97US-0059263.

PR 15-OCT-1997; 97US-0062125.

PR 17-OCT-1997; 97US-0062285.

PR 17-OCT-1997; 97US-0062287.

PR 24-OCT-1997; 97US-0062816.

PR 29-OCT-1997; 97US-0063704.

XX

PA (GENTH ) GENTTECH INC.

PI Botstein D, Goddard A, Gurney A, Hillan K, Lawrence DA;

PI Roy M, Wood WI;

XX

DR WPI: 1999-229532/19.

DR N-PSDB; AAX28432.

XX

PT Antibodies against specific proteins overexpressed in tumours

PT

XX

PS Example 1; Fig 12; 130pp: English.

XX

XX This sequence represents the EGF-like homologue PRO240.

CC The invention relates to antibodies (Ab) that bind to any of the

CC polypeptides (1) designated PRO187; PRO533; PRO214; PRO240; PRO211;

CC PRO230; PRO261; PRO246 or EBAF-2. The Ab, or other agents that inhibit

CC expression and/or activity of (1) are used: (1) to inhibit growth of

CC tumours; and (1i) as diagnostic/prognostic reagents for detection or

CC quantification of (1) in cells or tissues, by standard immunoassays, with

CC overexpression being indicative of cancer. For therapeutic use, the Ab

CC may be conjugated to a toxin, chemotherapeutic agent or radioisotope.

CC Genes expressing (1), many of which are growth factor homologues, are

CC overexpressed in some cases of cancer.

XX

Sequence 229 AA:

Query Match 98.8%; Score 959; DB 20; Length 229;  
 Best Local Similarity 98.9%; Pred. No. 9e-92;  
 Matches 175; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAPHGPSLTLLVPMWAAALLALGVERALALPEICTQCGSVQNSKVAFYCKTRELML 60  
 |||  
 DB 1 maphpgslltllvpwaaallalgyeralalpelctcpgsvqnlskvalfcktlxelm 60  
 QY 61 HARCCLNCKGTLTGLDQNSCLEDPGNFHOAHHTVTIIDQANPLKGDLANTEFGFTOLQ 120  
 |||  
 DB 61 haxccclnqkgtllgldqncslcdpgpnfhgahltvtiildqanplkgdiantfrgtlq 120  
 QY 121 TLILPQHVNCPEGGINAMWNTTTSYIDNQICOGKKNLCNNTGDPMECPENSGCVPDPG 177  
 |||  
 DB 121 tlllpqhvnpcpgginawntlttsyidnqicgqknlcntgdpemcpgnscvdpdp 177

## RESULT 11

AAG00157 standard; Protein: 147 AA.

AC AAG00157;  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein, SEQ ID NO: 4238.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 XX gene therapy; chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI03401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX

PA (GSEST ) GENSET.  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 DR N-PSDB; AAC00163.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 13; SEQ ID 4238; 71pp + CD-ROM; English.  
 XX

XX The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 CC

XX Sequence 147 AA;

Query Match 81.4%; Score 790; DB 21; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-74;

Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPHGPSLTLLVPMWAAALLALGVERALALPEICTQCGSVQNSKVAFYCKTRELML 60  
 |||  
 DB 1 maphpgslltllvpwaaallalgyeralalpelctcpgsvqnlskvalfcktlxelm 60  
 QY 61 HARCCLNCKGTLTGLDQNSCLEDPGNFHOAHHTVTIIDQANPLKGDLANTEFGFTOLQ 120  
 |||  
 DB 61 haxccclnqkgtllgldqncslcdpgpnfhgahltvtiildqanplkgdiantfrgtlq 120  
 QY 121 TLILPQHVNCPEGGINAMWNTTTSYIDNQ 147  
 |||  
 DB 121 tlllpqhvnpcpgginawntlttsyidnq 147

## RESULT 12

AAU04296 standard; Protein: 52 AA.

AAU04296;

DT 24-OCT-2001 (first entry)  
 XX  
 DE Transforming growth factor (TGF) alpha HI.  
 XX  
 KW Human; TGF alpha HI; transforming growth factor alpha HI; cancer;  
 XX diagnostic; therapeutic; immune disorder; multiple sclerosis;  
 XX systemic lupus erythematosus; human immunodeficiency virus; HIV;  
 KW hyperproliferative disorder; Gaucher's disease; cardiovascular disease;  
 KW Schmittler syndrome; Chaga's cardiomyopathy; coronary arteriosclerosis;  
 KW angiotensin disorder; corneal graft; neovascularisation; wound healing;  
 KW diabetic retinopathy; neurological disorder; Huntington's chorea;  
 KW Alzheimer's disease; Parkinson's disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200140251-A1.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PF 01-DEC-2000; 2000WO-US32745.  
 XX  
 PR 02-DEC-1999; 99US-0168387.  
 XX

PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Wei Y;  
 DR WPI: 2001-441480/47.  
 XX  
 PT Nucleic acid encoding human transforming growth factor alpha III  
 PT (TGFa), useful for preventing, diagnosing and/or treating e.g. Cancer  
 PT and Parkinson's disease -  
 XX  
 PS Disclosure: Fig 2: 302pp; English.  
 XX

XX The sequence represents the amino acid sequence of human transforming  
 CC growth factor (TGF) alpha HI. TGF alpha HI nucleic acid and protein  
 CC may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate polypeptide expression, for example immune  
 CC disorders (e.g. multiple sclerosis, systemic lupus erythematosus and  
 CC human immunodeficiency virus (HIV) infections), hyperproliferative  
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases  
 CC (e.g. Schmittler syndrome, Chaga's cardiomyopathy and coronary  
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 CC neovascularisation and diabetic retinopathy), neurological disorders  
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 CC infectious diseases and/or for promoting wound healing, regeneration  
 CC and/or chemotaxis (full details given in specification). Additionally,  
 CC the nucleic acid may be used to produce the secreted polypeptides, by  
 CC inserting the nucleic acids into a host cell and culturing the cell to  
 CC express the protein. It may also be used as a DNA probe in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acid

FN WO200042184-A1.  
XX  
XX  
PD 20-JUL-2000.  
XX  
PF 12-JAN-2000; 2000WO-US00742.  
XX  
PR 13-JAN-1999; 99US-0229598.  
XX  
PA (ZYMO ) ZYMOGENETICS INC.  
XX  
XX  
PI Piddington CS;  
XX  
XX  
DR WPI: 2000-476058/41.  
DR N-PSDB; AAA58786.  
XX  
XX  
PT New polypeptide, designated Zlrr3 for treating abnormal cell

PR	20-DEC-1999;	99NMO-US30911.
PR	16-DEC-1999;	99NMO-US30065.
PR	09-DEC-1999;	99NMO-US28565.
PR	02-DEC-1999;	99NMO-US28564.
PR	02-DEC-1999;	99NMO-US28551.
PR	02-DEC-1999;	99NMO-US28304.
PR	01-DEC-1999;	99NMO-US28631.
PR	01-DEC-1999;	99NMO-US28301.

PR	20-DEC-1999;	99NMO-US30911.
PR	16-DEC-1999;	99NMO-US30065.
PR	09-DEC-1999;	99NMO-US28565.
PR	02-DEC-1999;	99NMO-US28564.
PR	02-DEC-1999;	99NMO-US28551.
PR	02-DEC-1999;	99NMO-US28304.
PR	01-DEC-1999;	99NMO-US28631.
PR	01-DEC-1999;	99NMO-US28303.



PR 20-DEC-1999; 99WO-US30999.  
 PR 30-DEC-1999; 99WO-US31243.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 PR 11-FEB-2000; 2000WO-US03565.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 22-FEB-2000; 2000WO-US04414.  
 PR 24-FEB-2000; 2000WO-US04914.  
 PR 24-FEB-2000; 2000WO-US05004.  
 PR 01-MAR-2000; 2000WO-US05601.  
 PR 20-MAR-2000; 2000WO-US07377.  
 PR 21-MAR-2000; 2000WO-US07532.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 17-MAY-2000; 2000WO-US13705.  
 PR 22-MAY-2000; 2000WO-US14042.  
 PR 30-MAY-2000; 2000WO-US14941.  
 PR 02-JUN-2000; 2000WO-US15264.  
 PR 10-NOV-2000; 2000WO-US30873.  
 (GETH) GEMENTECH INC.  
 Baker RP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W, Gerritsen ME, Goddard A, Godowski PJ, Gunney AL, Sherwood S, Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
 WPI: 2001-408281/43.  
 N-PSDB: AAS21270.  
 DR Isolated, secretory and transmembrane PRO polypeptide used to detect  
 PT other PRO polypeptides, link bioactive molecules to cells expressing  
 PT lung, breast, prostate, cervical  
 PT lung, breast, prostate, cervical  
 PS Claim 12; Fig 54; 813pp; English.  
 XX AAU12172-AAU12446 represent novel human secretory and transmembrane  
 XX PRO polypeptides. The PRO polypeptides are useful to detect other  
 CC PRO polypeptides, to link bioactive molecules to cells expressing  
 CC PRO polypeptides, to modulate biological activities of cells expressing  
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,  
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO  
 CC polypeptide expression in a cell sample to that in a control sample.  
 CC Some of the 275 sequences are also useful to stimulate the release of  
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the  
 CC proliferation or differentiation of chondrocytes, the proliferation or  
 CC gene expression in pericyte cells, the release of proteoglycans from  
 CC cartilage, the proliferation of inner ear utricular supporting cells or  
 CC of T-lymphocytes, the release of a cytokine from peripheral blood  
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of  
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by  
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide  
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify  
 CC molecules involved in binding interactions. The polynucleotides encoding  
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,  
 CC transgenic or knock out animals and can be used in gene therapy.  
 XX  
 XX Sequence 281 AA:  
 SQ  
 Query Match 9.7%; Score 94.5; DB 22; Length 281;  
 Best Local Similarity 24.9%; Pred. No. 0.1;  
 Matches 53; Conservative 18; Mismatches 75; Indels 67; Gaps 10;

Db 124 hidshnpwlrirhpqafgqlmqrldlsyglafstlealegipgyltldqgnpwwc 183  
 Oy 138 ----NTTSYIDNQI--CQCGKMLCUNNTGDPDM 164  
 Db 184 gctmepllkwlrnirgrctadsqiaecrppcv 216  
 RESULT 15  
 AAR60176  
 ID AAR60176 standard; Protein; 385 AA.  
 XX  
 AC AAR60176;  
 DT 10-APR-1995 (first entry)  
 XX  
 DE Chimeric protein (R2:S) used in multivalent vaccine.  
 XX Chimeric; chimera; vaccine; multivalent; hepatitis B virus; HBV;  
 KW hepatitis; Japanese encephalitis virus; baculovirus.  
 XX Chimeric Hepatitis B virus.  
 OS Chimeric Japanese encephalitis virus.  
 PN JP06205672-A.  
 XX 26-JUL-1994.  
 PD  
 XX 19-MAR-1992; 92JP-0063699.  
 XX 19-MAR-1992; 92JP-0063699.  
 XX 19-MAR-1992; 92JP-0063699.  
 PA (JAPG) NIPPON ZEON KK.  
 PA (TOKS-) TOKYO SHINKAI KAGAKU SOGO KENKYUSHO ZH.  
 DR N-PSDB: AAG70157.  
 DR WPI: 1994-275516/34.  
 XX Prod. of chimeric proteins having antigenic sites from Japanese  
 PT encephalitis virus and hepatitis B virus surface antigens - also  
 PT recombinant baculovirus, useful as multivalent vaccine  
 XX disclosure; figures 9-12; 13pp; Japanese.  
 PS The chimeric protein comprises antigenic sites from Japanese  
 CC encephalitis virus and Hepatitis B virus surface antigens. The  
 CC protein may be used as a multivalent vaccine.  
 XX  
 XX Sequence 385 AA:  
 SQ  
 Query Match 8.7%; Score 84.5; DB 15; Length 385;  
 Best Local Similarity 24.7%; Pred. No. 1.7;  
 Matches 45; Conservative 23; Mismatches 61; Indels 53; Gaps 11;

Oy 24 GVEALALPEICTQCGSVONLSKVAFCYCKTRELMLHARCLNO-----KGTILGDLQ 78  
 Db 74 glhgala-----gal-----vdyssvklstgshlkrklmnpklkgtlygmet 119  
 Oy 79 NCSL---EDPGNFHQHTTIIIDQAN---PLK---GD---LANTFNGF----- 116  
 Db 120 ksfaknsadg-----hgtvlelsygsdgpcklsrtgdpapmenttsqfipllv 173  
 Oy 117 TQLOTLIPQHVNCPCGINAMNTITSYIDN-QICQCGKMLCUNNTGDPDMCPCNGSCVPDG 175  
 Db 174 lgaqflllrltllpqsldswtslnflgagapcpq-----nsqpslnhptscpic 228  
 Oy 176 PG 177  
 Db 229 PG 230



Thu Mar 28 09:21:42 2002

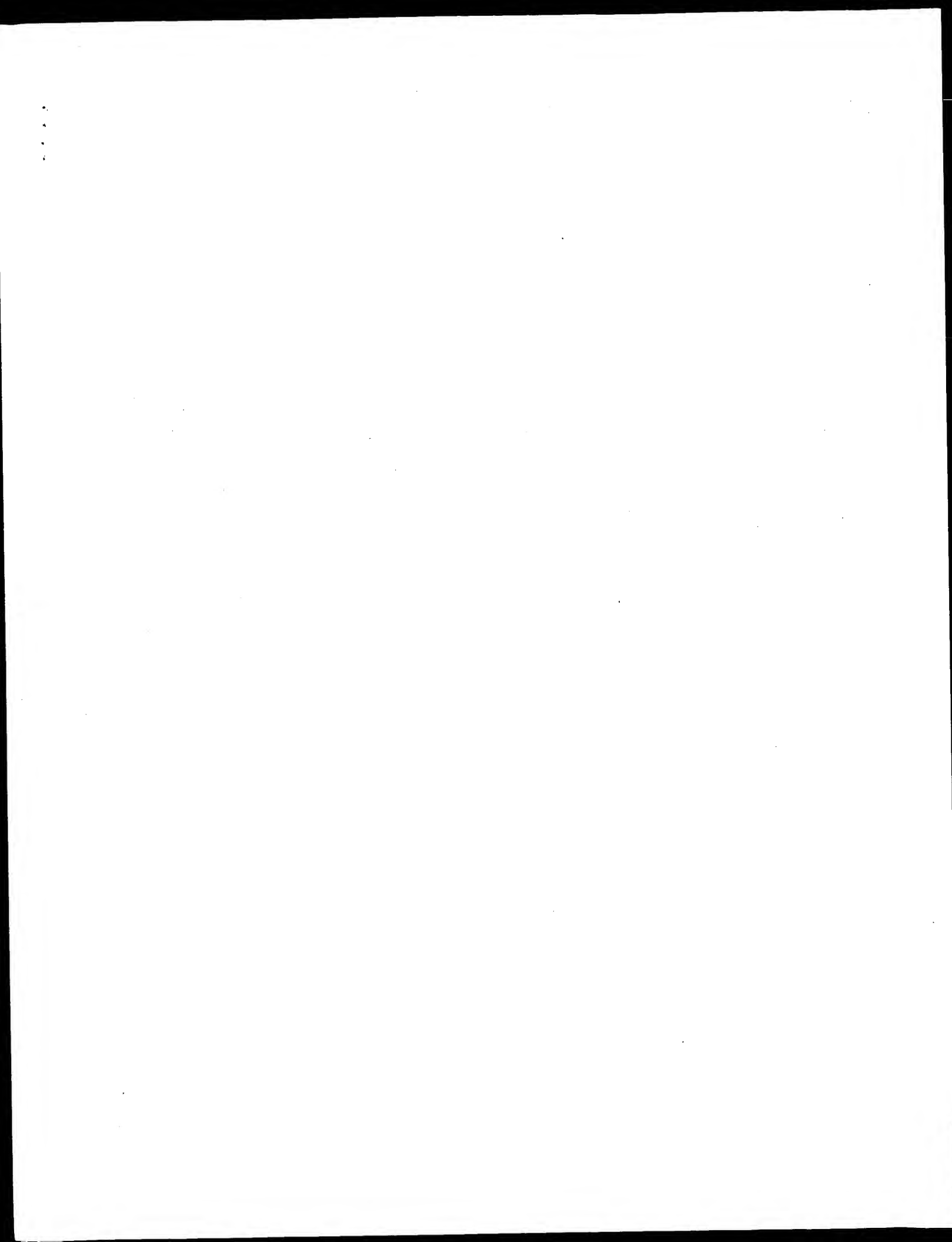
us-09-726-348-2\_copy\_1\_177\_1.rag

Page 11

Search completed: March 28, 2002, 08:58:42  
Job time: 5169 sec

0

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:15:53 ; Search time 59.64 Seconds

(Without alignments)  
219.835 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_1\_177

Perfect score: 177  
Sequence: 1 MAPHGPGLTTLVPMMAALL.....NTGDPEMPCENGSCVDPDPSG 177

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size: 0

Total number of hits satisfying chosen parameters: 237521

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database:

1: /SIDSL/gcgdata/geneseq/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/AA2001.DAT:\*

Pred: No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	4.0	23	22	Human transmembran
2	7	4.0	24	22	Maize hydroxyproli
3	6	3.4	9	22	HLA binding TADG-1
4	6	3.4	9	22	HLA binding TADG-1
5	6	3.4	9	22	HLA binding TADG-1
6	6	3.4	9	22	HLA binding TADG-1
7	6	3.4	9	22	HLA binding TADG-1
8	6	3.4	9	22	HLA binding TADG-1
9	6	3.4	9	22	HLA binding TADG-1
10	6	3.4	10	21	C. maltosa cytochr
11	6	3.4	12	15	Protonectarina pep

12	6	3.4	15	16	AA876703	Dendritic branched
13	6	3.4	15	22	AA811403	C. maltosa cytochr
14	6	3.4	18	22	AA836724	Peptide #10761 enc
15	6	3.4	19	21	AA822965	White shrimp multi
16	6	3.4	19	21	AA822965	White shrimp multi
17	6	3.4	19	22	AA864214	Grp78 peptide frag
18	6	3.4	20	20	AA809157	Peptide seq Id No:
19	6	3.4	20	21	AA822979	White shrimp multi
20	6	3.4	21	19	AA870528	White shrimp hydro
21	6	3.4	21	21	AA822980	White shrimp hydro
22	6	3.4	21	21	AA822986	White shrimp hydro
23	6	3.4	24	21	AA822988	White shrimp hydro
24	6	3.4	25	19	AA860083	N-terminal fragmen
25	6	3.4	25	20	AA814821	A. thaliana RAP2.7
26	6	3.4	25	21	AA822987	RAP2.7 linker regi
27	6	3.4	26	20	AA831494	Hepatitis GB virus
28	6	3.4	26	21	AA825853	Peptide #4659 enco
29	6	3.4	26	21	AA809081	Peptide #4753 enco
30	6	3.4	28	22	AA818225	Rice beta-glucanas
31	6	3.4	28	22	AA830716	2. may's partial we
32	6	3.4	28	22	AA805838	Cyclopropylbenzid
33	6	3.4	30	20	AA809284	Peptide antigen us
34	5	2.8	6	21	AA896772	Human erythropoiet
35	5	2.8	6	22	AA874917	Interleukin-1 rece
36	5	2.8	7	15	AA856275	Nuclear localisati
37	5	2.8	7	19	AA862062	Vasopressin (ADH)
38	5	2.8	7	19	AA853903	Vasopressin (ADH)
39	5	2.8	7	21	AA828277	Consensus peptide,
40	5	2.8	7	22	AA890878	Cytochrome express
41	5	2.8	7	22	AA890886	Human TSH receptor
42	5	2.8	7	22	AA890896	
43	5	2.8	8	14	AA842764	
44	5	2.8	8	15	AA847870	
45	5	2.8	8	16	AA873249	

## ALIGNMENTS

RESULT	1
ID	AA835198 standard; protein; 23 AA.
AC	AA835198;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Human transmembrane MUC3 fragment #4.
XX	
KW	Human: transmembrane MUC3; chromosome 7q22; colorectal cancer;
KW	Inflammatory bowel disease; chronic bronchitis; asthma; cystic fibrosis.
OS	Homo sapiens.
XX	
PN	WO200104152-A1.
XX	
PD	18-JAN-2001.
XX	
PF	13-JUL-2000; 2000WO-AU00846.
XX	
PR	13-JUL-1999; 99NZ-0336726.
XX	
PA	(MCGU/) MCGUCKIN M A.
PA	(WILL/) WILLIAMS S J.
XX	
PI	McGuckin MA, Williams SJ;
XX	
DR	WPI; 2001-138317/14.
XX	
PT	Novel transmembrane mucin 3 protein useful for producing
PT	anti-transmembrane mucin 3 antibodies useful in diagnosis and prognosis
PT	of colorectal cancer, inflammatory bowel disease and detecting
PT	transmembrane mucin 3 -

XX Claim 1; Page 32; 42pp; English.

CC The present invention provides the protein and coding sequences of the  
CC human transmembrane protein MUC3. The gene encoding this protein is found  
CC on human chromosome 7q22. The sequences can be used in the diagnosis of,  
CC detection of predisposition to and prognosis of inflammatory bowel  
CC disease, colorectal cancer, chronic bronchitis, asthma, cystic fibrosis  
CC and epithelial cancers, and in the identification of treatments for the  
CC same diseases. The present sequence is a fragment of the MUC3 protein.

XX Sequence 23 AA;

Query Match 4.0%; Score 7; DB 22; Length 23;  
Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 LILALGV 25  
Db 16 LILALGV 22

# RESULT 2

AAB66904 standard; Peptide: 24 AA.

AC AAB66904;

DT 12-APR-2001 (first entry)

DE Maize hydroxyproline-rich glycoprotein signal peptide.

KW Insecticide; transgenic plant; insect-resistance; signal peptide.

OS Zea mays.

PN WO200100841-A1.

PD 04-JAN-2001.

PE 23-JUN-2000; 2000WO-GB02457.

PR 29-JUN-1999; 99GB-0015215.

PR 23-DEC-1999; 99GB-0030536.

PA (ZENE) ZENECA LTD.

PI Griffin J, Carlile AJ, Cayley PJ, Mackay EA, Warner SAJ;

PI Vincent JL, Lee MD;

DR WPI; 2001-123015/13.

XX Novel insecticidal protein obtained from species of Paecilomyces for

PT controlling insects, and for insect-resistant transgenic plant

PS production

XX Disclosure; Pages 47-48; 72pp; English.

CC The present invention relates to novel insecticidal proteins obtained

CC from Paecilomyces sp. (see AAB66899 to AAB66901 and AAB66913). The

CC insecticidal proteins can be used to produce transgenic plants, which

CC are insect-resistant. Also, the insecticidal proteins are useful for

CC controlling insects by providing them at a locus where insects feed. The

CC present sequence is a signal peptide used in the present invention.

XX Sequence 24 AA;

Query Match 4.0%; Score 7; DB 22; Length 24;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AALLAL 23  
Db 7 aallal 13

# RESULT 3

AAU02236 standard; Peptide: 9 AA.

AC AAU02236;

DT 29-AUG-2001 (first entry)

DE HLA binding TADG-16 peptide #12.

KW Human; extracellular serine protease; tumour antigen derived gene-16;

KW TADG-16; ovarian carcinoma; breast cancer; lung cancer; colon cancer;

KW prostate cancer; HLA type.

OS Homo sapiens.

PN WO200127257-A1.

PD 19-APR-2001.

PE 13-OCT-2000; 2000WO-US28558.

PR 14-OCT-1999; 99US-0418527.

PA (UYAR) UNIV ARKANSAS.

PI O'Brien TJ, Underwood LJ, Shigemasa K;

DR WPI; 2001-273769/28.

XX New tumour antigen-derived gene-16 protein, useful for diagnosis and

PT treatment of ovarian, breast, lung, colon and prostate cancer -

PS Example 8; Page 52; 124pp; English.

CC AAU02225-AAU02384 represent TADG-16 peptides which are tested for

CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,

CC HLA A1, HLA B24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour

CC antigen derived gene-16 protein, TADG-16 (AAU02223), is a novel human

CC extracellular serine protease. TADG-16 is expressed in normal ovaries

CC and testes and in certain ovarian carcinomas. TADG-16 contains the

CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence

CC characteristic of the serine protease family. An antisense

CC oligonucleotide having a complementary sequence to the TADG-16 nucleic

CC acid is useful for treating various cancers, including ovarian, breast,

CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and

CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.

CC TADG-16 protein or its fragments are useful for vaccinating an individual

CC against TADG-16.

XX Sequence 9 AA;

Query Match 3.4%; Score 6; DB 22; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 ALLIAL 23  
Db 1 allial 6

# RESULT 4

AAU02263 standard; Peptide: 9 AA.

AC AAU02263;

[illegible]

DR WPI: 2001-273769/28.

XX New tumour antigen-derived gene-16 protein, useful for diagnosis and  
PT treatment of ovarian, breast, lung, colon and prostate cancer -  
XX

PS Example 8; Page 54; 124pp; English.

CC AAU02225-AAU02384 represent TADG-16 peptides which are tested for  
CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,  
CC HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour  
CC antigen derived gene-16 protein, TADG-16 (AAU02223), is a novel human  
CC extracellular serine protease. TADG-16 is expressed in normal ovaries  
CC and testes and in certain ovarian carcinomas. TADG-16 contains the  
CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
CC characteristic of the serine protease family. An antisense  
CC oligonucleotide having a complementary sequence to the TADG-16 nucleic  
CC acid is useful for treating various cancers, including ovarian, breast,  
CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
CC TADG-16 protein or its fragments are useful for vaccinating an individual  
CC against TADG-16.

XX Sequence 9 AA;

Query Match 3.4%; Score 6; DB 22; Length 9;

Best Local Similarity 100.0%; Pred. No. 4.3e+05; Mismatches 0; Gaps 0;

Matches 6; Conservative 0; Indels 0;

OY 18 ALLAL 23

DB 2 allal 7

RESULT 7

ID AAU02315 standard; Peptide; 9 AA.

XX AAU02315;

DT 29-AUG-2001 (first entry)

XX HLA binding TADG-16 peptide #91.

XX Human; extracellular serine protease; tumour antigen derived gene-16;

KW TADG-16; ovarian carcinoma; breast cancer; lung cancer; colon cancer;

XX prostate cancer; HLA type.

XX Homo sapiens.

XX WO200127257-A1.

XX 19-APR-2001.

XX 13-OCT-2000; 2000MO-US28558.

XX 14-OCT-1999; 99US-0418527.

XX (UYAR-) UNIV ARKANSAS.

XX O'Brien TJ, Underwood LJ, Shigemasa K;

XX WPI: 2001-273769/28.

XX New tumour antigen-derived gene-16 protein, useful for diagnosis and

PT treatment of ovarian, breast, lung, colon and prostate cancer -

XX Example 8; Page 54; 124pp; English.

CC AAU02225-AAU02384 represent TADG-16 peptides which are tested for  
CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,  
CC HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour  
CC antigen derived gene-16 protein, TADG-16 (AAU02223), is a novel human  
CC antigen derived gene-16 protein, TADG-16 (AAU02223), is a novel human

CC extracellular serine protease. TADG-16 is expressed in normal ovaries  
CC and testes and in certain ovarian carcinomas. TADG-16 contains the  
CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
CC characteristic of the serine protease family. An antisense  
CC oligonucleotide having a complementary sequence to the TADG-16 nucleic  
CC acid is useful for treating various cancers, including ovarian, breast,  
CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
CC TADG-16 protein or its fragments are useful for vaccinating an individual  
CC against TADG-16.

XX Sequence 9 AA;

Query Match 3.4%; Score 6; DB 22; Length 9;

Best Local Similarity 100.0%; Pred. No. 4.3e+05; Mismatches 0; Gaps 0;

Matches 6; Conservative 0; Indels 0;

OY 18 ALLAL 23

DB 2 allal 7

RESULT 8

ID AAU02334 standard; Peptide; 9 AA.

XX AAU02334;

DT 29-AUG-2001 (first entry)

XX HLA binding TADG-16 peptide #110.

XX Human; extracellular serine protease; tumour antigen derived gene-16;

KW TADG-16; ovarian carcinoma; breast cancer; lung cancer; colon cancer;

XX prostate cancer; HLA type.

XX Homo sapiens.

XX WO200127257-A1.

XX 19-APR-2001.

XX 13-OCT-2000; 2000MO-US28558.

XX 14-OCT-1999; 99US-0418527.

XX (UYAR-) UNIV ARKANSAS.

XX O'Brien TJ, Underwood LJ, Shigemasa K;

XX WPI: 2001-273769/28.

XX New tumour antigen-derived gene-16 protein, useful for diagnosis and

PT treatment of ovarian, breast, lung, colon and prostate cancer -

XX Example 8; Page 55; 124pp; English.

CC AAU02225-AAU02384 represent TADG-16 peptides which are tested for  
CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,  
CC HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour  
CC antigen derived gene-16 protein, TADG-16 (AAU02223), is a novel human  
CC extracellular serine protease. TADG-16 is expressed in normal ovaries  
CC and testes and in certain ovarian carcinomas. TADG-16 contains the  
CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
CC characteristic of the serine protease family. An antisense  
CC oligonucleotide having a complementary sequence to the TADG-16 nucleic  
CC acid is useful for treating various cancers, including ovarian, breast,  
CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
CC TADG-16 protein or its fragments are useful for vaccinating an individual  
CC against TADG-16.

SQ Sequence 9 AA;

Query Match 3.4%; Score 6; DB 22; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 ALLIAL 23  
 |||||  
 Db 2 alllal 7

RESULT 9

AAU02351

ID AAU02351 standard; Peptide; 9 AA.

XX AC AAU02351;

DT 29-NOV-2001 (first entry)

DE HLA binding TADG-16 peptide #127.

XX Human; extracellular serine protease; tumour antigen derived gene-16;

KW TADG-16; ovarian carcinoma; breast cancer; lung cancer; colon cancer;

XX prostate cancer; HLA type.

OS Homo sapiens.

XX W0200127257-A1.

PD 19-APR-2001.

PF 13-OCT-2000; 2000MO-US28558.

PR 14-OCT-1999; 990S-0418527.

XX (UYAR-) UNITV ARKANSAS.

XX O'Brien TJ, Underwood LJ, Shigemasa K;

DR WPI; 2001-273769/28.

PT New tumour antigen-derived gene-16 protein, useful for diagnosis and  
 treatment of ovarian, breast, lung, colon and prostate cancer -

PS Example 8; Page 55; 124pp; English.

CC AAU02225-AAU02384 represent TADG-16 peptides which are tested for  
 CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,  
 CC HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour  
 CC antigen derived gene-16 protein, TADG-16 (AAU02223), is a novel human  
 CC extracellular serine protease. TADG-16 is expressed in normal ovaries  
 CC and testes and in certain ovarian carcinomas. TADG-16 contains the  
 CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
 CC characteristic of the serine protease family. An antisense  
 CC oligonucleotide having a complementary sequence to the TADG-16 nucleic  
 CC acid is useful for treating various cancers, including ovarian, breast,  
 CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
 CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
 CC TADG-16 protein or its fragments are useful for vaccinating an individual  
 CC against TADG-16.

SQ Sequence 9 AA;

Query Match 3.4%; Score 6; DB 22; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 ALLIAL 23  
 |||||  
 Db 4 alllal 9

RESULT 10

AAB11404

ID AAB11404 standard; peptide; 10 AA.

XX AC AAB11404;

DT 22-FEB-2001 (first entry)

DE C. maltosa cytochrome b5 peptide fragment T54.

XX Cytochrome b5; alkane metabolism; oxidation; long-chain alkyl compound;

XX long-chain dicarboxylic acid.

XX Candida maltosa.

XX W0200065061-A2.

XX 02-NOV-2000.

XX 18-APR-2000; 2000MO-DE01246.

XX 24-APR-1999; 99DE-1018763.

XX (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

XX Schunck W, Chernogolov A;

XX WPI; 2000-679674/66.

PT Nucleic acid sequences from alkane metabolizing Candida yeast, encoding  
 PT cytochrome b5 and used for the oxidation of long chain alkyl compounds  
 PT and for the production of long chain dicarboxylic acids -

PS Example 1.2; Page 6; 27pp; German.

CC This invention describes novel nucleic acid sequences from alkane  
 CC metabolizing Candida yeasts, encoding cytochrome b5-polypeptide, its  
 CC fragments, variants and mutations. The nucleic acids and polypeptides  
 CC are used for the oxidation of long-chain alkyl compounds with at least  
 CC 10C, and for the production of long-chain dicarboxylic acids by oxidizing  
 CC n-alkanes and fatty acids with at least 10C.

SQ Sequence 10 AA;

Query Match 3.4%; Score 6; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 140 ITSYID 145  
 |||||  
 Db 4 itsyid 9

RESULT 11

AAR45452

ID AAR45452 standard; peptide; 12 AA.

XX AC AAR45452;

DT 11-JUL-1994 (first entry)

DE Protonecarina peptide p1 with histamine releasing activity.

XX Hunting wasp; poison sac; histamine release; secretion; allergy;

XX inflammation; mast cell.

XX Protonecarina sylvestrae.

OS Key Location/Qualifiers  
 FH Modified-site 12  
 FT /note- "in amide form"

XX JP05331194-A.  
 XX 14-DEC-1993.  
 PD 03-JUN-1992: 92JP-0142299.  
 XX 03-JUN-1992: 92JP-0142299.  
 XX 03-JUN-1992: 92JP-0142299.  
 XX (DAIL) DAICEL CHEM IND LTD.  
 XX WPI, 1994-022916/03.  
 DR Bioactive peptide obtd. from hunting wasp - used for development  
 XX of research reagents and drugs for allergy and inflammation  
 PT related to mast cell  
 PS Claim 1; Page 8; 10pp; Japanese.  
 CC Protocetaria peptide P-1 was isolated from the poison sac of the  
 CC hunting wasp P. sylvestris. The peptide has a high histamine releasing  
 CC activity on rat abdominal mast cells and will be useful for the  
 CC development of research reagents and drugs for allergy and  
 CC inflammation related to mast cells. See also AAR45453 and AAR45454.  
 XX  
 SQ Sequence 12 AA:

Query Match 3.4%; Score 6; DB 15; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 70 GTTGL 75  
 DB 3 GTTGL 8

RESULT 12  
 AAR76703  
 ID AAR76703 standard; peptide; 15 AA.  
 XX  
 AC AAR76703;  
 XX  
 DT 31-JAN-1996 (first entry)  
 XX  
 DE Dendritic branched signal peptide; DBP-3.  
 XX  
 KW Dendritic; branched; membrane; signal peptide; cerebral; hormones;  
 KW polylysine core; drug delivery.  
 XX  
 OS Synthetic.  
 XX  
 XX  
 FH Key  
 FT Modified-site  
 FT Location/Qualifiers  
 FT /label= branch-site  
 FT /note= "omega amino group in the sidechain of  
 FT Lys(5) forms a peptide bond with the  
 FT C-terminus of the tetrapeptide  
 FT Arg-Leu-Leu-Leu"  
 FT Modified-site  
 FT /label= branch-site  
 FT /note= "omega amino group in the sidechain of  
 FT Lys(6) forms a peptide bond with the  
 FT C-terminus of the peptide  
 FT Arg-Leu-Leu-Lys. Omega amino group  
 FT of Lys of adjoining peptide forms a  
 FT tetrapeptide bond with the C-terminus of the  
 FT tetrapeptide Arg-Leu-Leu-Leu"  
 FT Modified-site  
 FT /label= branch-site  
 FT /note= "omega amino group in the side chain of  
 FT Lys(7) forms a peptide bond with the  
 FT carboxyl group of a Lys residue analogous

FT to Lys(6)"  
 XX JP07126287-A.  
 XX 16-MAY-1995.  
 PD 05-NOV-1993: 93JP-0276912.  
 XX 05-NOV-1993: 93JP-0276912.  
 XX 05-NOV-1993: 93JP-0276912.  
 XX (FUGI/) FUGI T.  
 XX WPI, 1995-212954/28.  
 DR Membrane recognising signal peptide - useful for aiding penetration  
 XX of drugs into cells, partic. hormones into cerebral cells.  
 PT  
 PS Disclosure: Fig 1; 4pp; Japanese.  
 CC AAR76703 is a polylysine core peptide of the formula (H-Arg-Leu-Leu-  
 CC Leu)-8-Lys4-Lys2-Ala-Ala-Leu-Leu-Leu-Arg-Arg). It is a  
 CC signal peptide which when conjugated to drugs promotes their  
 CC passage through the cell membrane and uptake into cells.  
 XX  
 SQ Sequence 15 AA:

Query Match 3.4%; Score 6; DB 16; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 AAALL 21  
 DB 8 aaall 13

RESULT 13  
 AAB11403  
 ID AAB11403 standard; peptide; 15 AA.  
 XX  
 AC AAB11403;  
 XX  
 DT 22-FEB-2001 (first entry)  
 XX  
 DE C. maltosa cytochrome b5 peptide fragment T53.  
 XX  
 DE Cytochrome b5; alkane metabolism; oxidation; long-chain alkyl compound;  
 KW long-chain dicarboxylic acid.  
 KW Candida maltosa.  
 XX  
 OS WO200065061-A2.  
 XX  
 PN 02-NOV-2000.  
 XX  
 PD 18-APR-2000; 2000WO-DE01246.  
 PF 24-APR-1999: 99DE-1018763.  
 XX  
 PR (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.  
 PA Schunck W, Chernogolov A;  
 PI WPI, 2000-679674/66.  
 DR Nucleic acid sequences from alkane metabolizing Candida yeast, encoding  
 XX cytochrome b5 and used for the oxidation of long chain alkyl compounds  
 PT and for the production of long chain dicarboxylic acids -  
 PT  
 PS Example 1.2; Page 6; 27pp; German.  
 XX This invention describes novel nucleic acid sequences from alkane  
 CC metabolizing Candida yeasts, encoding cytochrome b5-polypeptide, its



CC fragments, variants and mutations. The nucleic acids and polypeptides  
 CC are used for the oxidation of long-chain alkyl compounds with at least  
 CC 10C, and for the production of long-chain dicarboxylic acids by oxidizing  
 CC n-alkanes and fatty acids with at least 10C.  
 SO Sequence 15 AA;

Query Match 3.4%; Score 6; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 140 ITSYID 145  
 Db 4 Itsyid 9

RESULT 14  
 AAM36724  
 ID AAM36724 standard; Protein; 18 AA.  
 XX  
 AC AAM36724;

DE 17-OCT-2001 (first entry)

Peptide #10761 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236559.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

DR Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 27; SEQ ID No 36993; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SNP).

XX see A131315-A131546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders.

SO Sequence 18 AA;

Db 8 lllalg 13

RESULT 15  
 AAB22969  
 ID AAB22969 standard; peptide; 19 AA.  
 XX  
 AC AAB22969;

DE 10-JAN-2001 (first entry)

White shrimp multifunctional hydrolase isoform p5.1a signal peptide.

XX Multifunctional hydrolase isoform; multifunctional protein;

XX white shrimp; protease activity; chymotrypsin; trypsin; collagenase;

XX elastase; wound healing; corneal ulcer; internal trauma;

XX internal surgical wound; nerve; tendon; sheath;

XX adhesion formation inhibition; opthalmological; vulnerary.

XX Penaeus vannamei.

OS Penaeus vannamei.

PN WO200049991-A2.

PD 31-AUG-2000.

PF 29-JUN-1999; 99WO-US14751.

PR 23-FEB-1999; 99US-0256484.

PR (PHAL-) PHARMACON MEDICAL INC.

XX Franklin R, Cowling D, Hubbell JA;

XX WPI; 2000-SF7120/55.

XX Use of microgel comprising crosslinked polyanionic polymer and

XX optionally proteolytic enzyme in treatment of corneal wounds and

XX ulcers and abrasions and internal trauma e.g. surgical wounds and for

XX treating implants to reduce adhesions -

XX Disclosure; Page 26; 85bp; English.

XX The invention relates to use of a microgel comprising a crosslinked

XX polyanionic polymer in compositions for treatment of an area affected by

XX a trauma selected from corneal wounds and internal trauma. The

XX compositions further comprise a multifunctional protease having

XX activities selected from two of chymotrypsin activity, trypsin activity,

XX collagenase activity and elastase activity. In particular, the protease

XX may be one of six isoforms of white shrimp (Penaeus vannamei)

XX comprising one of the peptide sequences given in AAB22963-B22967.

XX (Including infected ulcers), or abrasions or a chemical or physical

XX insult to the cornea that are likely to give rise to a corneal ulcer.

XX They are also used to treat internal traumas selected from an

XX internal surgical wound, or a trauma to a membrane that covers either

XX an internal organ or tissue or the cavity in which one or more internal

XX organs or tissues reside, with the membrane being selected from the

XX peritoneum, the pericardium, the epicardium and the pleura and a

XX meninges, the internal trauma is to a tendon, tendon sheath, a nerve

XX or a nerve sheath, where the trauma is susceptible to giving rise to

XX adhesions. The amount of microgel administered is sufficient to prevent

XX or reduce formation or reformation of adhesions, and the composition can

XX also be used to treat an implant to reduce formation of adhesions.

XX Sequences AAB22968-B22980 and AAB22985-B22988 represent fragments or

XX derived peptides of white shrimp multifunctional hydrolase isoforms

XX referred to in the disclosure of the invention.

SO Sequence 19 AA;

Query Match 3.4%; Score 6; DB 21; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 63;

OY 19 lllalg 24  
 |||||

Thu Mar 28 09:21:39 2002

us-09-726-348-2\_copy\_1\_177.rag

Matches	6;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Oy	18	ALLAL	23						
Db	7	alllal	12						

Search completed: March 28, 2002, 09:15:53  
Job time: 852 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 08:58:42 ; Search time 105.47 seconds  
(without alignments)  
36.520 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_126\_177

Perfect score: 309

Sequence: 1 QHVNCPGGINAMWITTSYID.....NTGDPENCPENGSCVPDPCG 52

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDSL1/gcgdata/geneseq/AA1980.DAT:\*  
2: /SIDSL1/gcgdata/geneseq/AA1981.DAT:\*  
3: /SIDSL1/gcgdata/geneseq/AA1982.DAT:\*  
4: /SIDSL1/gcgdata/geneseq/AA1983.DAT:\*  
5: /SIDSL1/gcgdata/geneseq/AA1984.DAT:\*  
6: /SIDSL1/gcgdata/geneseq/AA1985.DAT:\*  
7: /SIDSL1/gcgdata/geneseq/AA1986.DAT:\*  
8: /SIDSL1/gcgdata/geneseq/AA1987.DAT:\*  
9: /SIDSL1/gcgdata/geneseq/AA1988.DAT:\*  
10: /SIDSL1/gcgdata/geneseq/AA1989.DAT:\*  
11: /SIDSL1/gcgdata/geneseq/AA1990.DAT:\*  
12: /SIDSL1/gcgdata/geneseq/AA1991.DAT:\*  
13: /SIDSL1/gcgdata/geneseq/AA1992.DAT:\*  
14: /SIDSL1/gcgdata/geneseq/AA1993.DAT:\*  
15: /SIDSL1/gcgdata/geneseq/AA1994.DAT:\*  
16: /SIDSL1/gcgdata/geneseq/AA1995.DAT:\*  
17: /SIDSL1/gcgdata/geneseq/AA1996.DAT:\*  
18: /SIDSL1/gcgdata/geneseq/AA1997.DAT:\*  
19: /SIDSL1/gcgdata/geneseq/AA1998.DAT:\*  
20: /SIDSL1/gcgdata/geneseq/AA1999.DAT:\*  
21: /SIDSL1/gcgdata/geneseq/AA2000.DAT:\*  
22: /SIDSL1/gcgdata/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	309	100.0	229	18	AAW27087
2	309	100.0	229	20	AAW13944
3	309	100.0	229	20	AAW105282
4	309	100.0	229	21	AAW33419
5	309	100.0	229	21	AAW8570
6	309	100.0	229	22	AAW04295
7	309	100.0	229	22	AAW20112
8	309	100.0	229	22	AAW68595
9	309	100.0	229	22	AAW50951
10	309	100.0	342	22	AAW48068
11	163	52.8	52	22	AAW04296

12	128	41.4	147	21	AAW00157	Human secreted pro
13	77	24.9	296	20	AAW27073	L. binastus plasmi
14	67	21.7	2019	16	AAW67913	Cardiac sodium cha
15	67	21.7	2020	11	AAW06584	Cardiac sodium cha
16	65.5	21.2	1050	22	AAW66267	Human TANGO 272 SE
17	63.5	20.6	880	21	AAW01249	Human EMR1 hormone
18	63.5	20.6	880	22	AAW1869	Human EMR1 hormone
19	63	20.4	2476	20	AAW67738	Pig p105 zona pell
20	62.5	20.2	1964	20	AAW95557	Mus musculus notch
21	62.5	20.2	2471	20	AAW06816	Human Notch2 (hum
22	61.5	19.9	568	20	AAW35121	Amino acid sequenc
23	61	19.7	321	19	AAW53243	Mus musculus vascu
24	61	19.7	337	20	AAW08286	Murine c-Fos induc
25	61	19.7	358	18	AAW14992	Mus musculus vascu
26	61	19.7	358	19	AAW53242	Mus musculus vascu
27	61	19.7	358	19	AAW44295	Mus musculus vascu
28	61	19.7	658	22	AAW06934	Mouse vascular end
29	61	19.7	802	20	AAW41710	Human PRO618 prote
30	61	19.7	802	21	AAW44266	Human PRO618 (UN03
31	61	19.7	802	21	AAW24052	Human PRO618 prote
32	61	19.7	802	22	AAW06933	Human membrane-ty
33	61	19.7	4655	17	AAW97209	Human membrane-ty
34	61	19.7	4655	17	AAW97209	Human parathyroid
35	61	19.7	4655	17	AAW97210	Human parathyroid
36	61	19.7	4655	19	AAW43311	Human parathyroid
37	61	19.7	4655	19	AAW43312	Human parathyroid
38	61	19.7	4655	19	AAW43312	Human parathyroid
39	61	19.7	4655	19	AAW43313	Human parathyroid
40	61	19.7	4655	19	AAW43314	Human parathyroid
41	60.5	19.6	497	18	AAW13106	Human placental ca
42	60.5	19.6	497	18	AAW11474	Human placental ca
43	60.5	19.6	2005	22	AAW96676	Human placental ca
44	60.5	19.6	2005	22	AAW96677	Human placental ca
45	60	19.4	1951	22	AAW96678	Human placental ca

## ALIGNMENTS

RESULT 1	
ID	AAW27087 standard; Protein: 229 AA.
AC	AAW27087;
XX	
DT	28-JAN-1998 (first entry)
XX	
DE	Human transforming growth factor alpha H11.
XX	
KW	human transforming growth factor; TGF; TGF-alpha-H11; angiogenesis;
KW	embryogenesis; ocular disorder; kidney disorder; liver disorder;
KW	neuronal disorder; alopecia; inflammation.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	Peptide
FT	1..25
FT	/label= signal_peptide
FT	/note= "putative"
FT	26..229
FT	/label= mature_protein
FT	1..177
FT	/label= soluble_portion_of_protein
FT	126..177
FT	/label= active_site
FT	178..204
FT	/label= transmembrane_portion
FT	/note= "putative"
XX	
PN	MO9725349-A1.
XX	
PD	17-JUL-1997.
XX	

PF 04-JAN-1996; 96WO-US00149.  
 XX  
 PR 04-JAN-1996; 96WO-US00149.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Wei Y;  
 DR WPI: 1997-372817/34.  
 DR N-PSDB: AAT85082.  
 XX  
 PT New human transforming growth factor-alpha homologue - used for  
 PT developing products for treating e.g. neurological disorders, kidney  
 PT and liver disorders, tumours, wounds, hair loss or skin disorders  
 XX  
 PS Claim 15; Page 47; 63pp; English.  
 CC This protein has been putatively identified as a human transforming  
 CC growth factor (TGF) alpha analogue, TGF-alpha-Hill. The protein can  
 CC stimulate angiogenesis, embryogenesis, cell differentiation and function.  
 CC It can be used for therapeutic purposes for restoration or enhancement of  
 CC neurological functions diminished as a result of trauma or other damaging  
 CC pathological functions such as AIDS dementia and senile dementia, to treat  
 CC disorders, e.g. corneal inflammation, to destroy target cells, to treat  
 CC tumours, kidney or liver disorders or to treat wounds, burns or ulcers.  
 CC The polypeptide can also be used in the modulation of angiogenesis, bone  
 CC resorption, immune response, and synaptic and neuronal effector  
 CC functions, or the arachidonic acid cascade. It can also be used in  
 CC applications related to terminal differentiation e.g. in  
 CC hyperproliferative disorders such as inflammation or psoriasis and for  
 CC alopecia, hair loss or other skin conditions which affect hair follicular  
 CC development. Antagonists to TGF-alpha-Hill can be used for treating  
 CC tumours or skin disorders such as psoriasis. The products can also  
 CC be used for diagnosis and detection of the above disorders.  
 XX  
 SQ Sequence 229 AA;

Query Match 100.0%; Score 309; DB 18; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-25;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GHVNCBGGINAWNTTTSYIDNOICOGKNCNNTGDEMPENGSCVPDGP 52  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 126 ghvncpgginawntlttsyidnqicgqknlcmntgdepengscvpdgp 177

## RESULT 2

AAV13944 standard; Protein: 229 AA.

ID AAV13944 standard; Protein: 229 AA.  
 AC AAV13944;  
 XX  
 DE 14-JUL-1999 (first entry)  
 XX  
 DE Human transmembrane protein, HP10435.  
 XX  
 KW Transmembrane protein; human; cell membrane; proliferation; diagnosis;  
 KW cell differentiation; carcinostatic agent; probe; gene therapy;  
 KW signal transduction; apoptosis; inhibitor;  
 KW phosphatidylethanolamine N-methyltransferase.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9918203-A2.  
 XX  
 PD 15-APR-1999.  
 XX  
 PF 05-OCT-1998; 98WO-JP04475.  
 XX  
 PR 08-OCT-1997; 97JP-0276271.  
 XX  
 PA (PROT-) PROTEGENE INC.

PA (SAGA) SAGAMI CHEM RES CENT.  
 XX  
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
 XX  
 DR WPI: 1999-277268/23.  
 DR N-PSDB: AAX36812, AAX36813.  
 XX  
 PT Human transmembrane proteins and nucleotide sequences  
 XX  
 PS Claim 1; Page 90-91; 139pp; English.  
 CC This sequence is a human transmembrane protein of the invention.  
 CC All of the proteins exist in the cell membrane, so are considered to be  
 CC proteins controlling the proliferation and differentiation of the cells.  
 CC They may be useful as carcinostatic agents or as antigens for preparing  
 CC antibodies against the proteins. The cDNAs can be used as probes for  
 CC gene diagnosis and gene sources for gene therapy, as well as for  
 CC large-scale expression of the proteins. The HP01498 (see AAV13939)  
 CC protein may be associated with signal transduction associated with  
 CC (see AAV13943) protein can be used to treat diseases associated with  
 CC phosphatidylethanolamine N-methyltransferase. The proteins are  
 CC identified by the presence of a hydrophobic transmembrane region,  
 CC knowledge of the protein function is not required, as in e.g. methods of  
 CC expression cloning.  
 XX  
 SQ Sequence 229 AA;

Query Match 100.0%; Score 309; DB 20; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-25;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GHVNCBGGINAWNTTTSYIDNOICOGKNCNNTGDEMPENGSCVPDGP 52  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 126 ghvncpgginawntlttsyidnqicgqknlcmntgdepengscvpdgp 177

## RESULT 3

AAV05282 standard; Protein: 229 AA.

ID AAV05282 standard; Protein: 229 AA.  
 AC AAV05282;  
 XX  
 DE 22-JUN-1999 (first entry)  
 XX  
 DE EGF-like homologue PRO240.  
 XX  
 KW Antibody; PRO187; PRO533; PRO214; PRO240; PRO261; PRO246;  
 KW EBAF-2; inhibitor; tumour growth; cancer; EGF-like homologue;  
 KW EGF-8 homologue.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9914327-A2.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 10-SEP-1998; 98WO-US18824.  
 XX  
 PR 25-NOV-1997; 97US-0066840.  
 PR 17-SEP-1997; 97US-0059114.  
 PR 17-SEP-1997; 97US-0059117.  
 PR 18-SEP-1997; 97US-0059263.  
 PR 15-OCT-1997; 97US-0062125.  
 PR 17-OCT-1997; 97US-0062285.  
 PR 17-OCT-1997; 97US-0062287.  
 PR 24-OCT-1997; 97US-0062816.  
 PR 29-OCT-1997; 97US-0063704.  
 XX  
 PA (GETH) GENENTECH INC.  
 XX  
 PI Bolstein D, Goddard A, Gurney A, Hillan K, Lawrence DA;

PI Roy M, Wood WI;  
 XX WPI: 1999-229532/19.  
 DR N-PSDB: AAX28432.  
 XX  
 PT Antibodies against specific proteins overexpressed in tumours  
 XX  
 PS Example 1: Fig 12: 130pp: English.  
 XX  
 CC This sequence represents the EGF-like homologue PRO240.  
 CC The invention relates to antibodies (Ab) that bind to any of the  
 CC polypeptides (I) designated PRO187; PRO533; PRO214; PRO240; PRO211;  
 CC PRO230; PRO261; PRO246 or EBAR-2. The Ab, or other agents that inhibit  
 CC expression and/or activity of (I) are used: (i) to inhibit growth of  
 CC tumours; and (ii) as diagnostic/prognostic reagents for detection or  
 CC quantification of (I) in cells or tissues, by standard immunoassays, with  
 CC overexpression being indicative of cancer. For therapeutic use, the Ab  
 CC may be conjugated to a toxin, chemotherapeutic agent or radioisotope.  
 CC Genes expressing (I), many of which are growth factor homologues, are  
 CC overexpressed in some cases of cancer.  
 CC  
 XX Sequence 229 AA;  
 SQ  
 Query Match 100.0%; Score 309; DB 20; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-25;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GHVNCPEGINAWNTTYSYIDNOICGCKNLGNMGDPENSCVDPGCG 52  
 DB 126 ghvncpgginawnttysyidngicgcknlcnmgdpemscvdpdpg 177  
 RESULT 4  
 AAB33419 standard; Protein: 229 AA.  
 XX AAB33419;  
 AC  
 XX  
 DT 29-JAN-2001 (first entry)  
 XX  
 DE Human PRO240 protein UNQ214 SEQ ID NO:26.  
 XX  
 KW Human; immune related disease; diagnosis; antiinflammatory; cardiac;  
 KW dermatological; antiarthritic; antirheumatic; immunosuppressive;  
 KW haemostatic; antithyroid; antidiabetic; neutrotropic; neuroprotective;  
 KW antinaeumatic; hepatotropic; virucide; antipsoriatic; antiallergic;  
 KW antiautlumatic; systemic lupus erythematosus; rheumatoid arthritis;  
 KW idiopathic inflammatory myopathy; Sjogren's syndrome; sarcoidosis;  
 KW autoimmune thrombocytopenia; autoimmune haemolytic anaemia; diabetes mellitus;  
 KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;  
 KW demyelinating disease; hepatobiliary disease; Whipple's disease;  
 KW inflammatory bowel disease; gluten-sensitive enteropathy;  
 KW autoimmune disease; immune-mediated skin disease; allergic disease;  
 KW immunological disease; transplantation associated disease;  
 KW graft rejection; graft-versus-host disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200053758-A2.  
 PD 14-SEP-2000.  
 XX  
 XX 02-MAR-2000: 2000WO-US05841.  
 PF 08-MAR-1999: 99WO-US05028.  
 PR 10-MAR-1999: 99US-0123618.  
 PR 12-MAR-1999: 99US-0123957.  
 PR 23-MAR-1999: 99US-0125775.  
 PR 12-APR-1999: 99US-0128849.  
 PR 20-APR-1999: 99WO-US08615.  
 PR 28-APR-1999: 99US-0131445.

PR 04-MAY-1999: 99US-0132371.  
 PR 14-MAY-1999: 99US-0134287.  
 PR 02-JUN-1999: 99WO-US12252.  
 PR 23-JUN-1999: 99US-0141037.  
 PR 20-JUL-1999: 99US-0144738.  
 PR 26-JUL-1999: 99US-0145698.  
 PR 28-JUL-1999: 99US-0146222.  
 PR 01-SEP-1999: 99WO-US20111.  
 PR 08-SEP-1999: 99WO-US20594.  
 PR 13-SEP-1999: 99WO-US20944.  
 PR 15-SEP-1999: 99WO-US21090.  
 PR 15-SEP-1999: 99WO-US21547.  
 PR 05-OCT-1999: 99WO-US23089.  
 PR 29-OCT-1999: 99US-0162506.  
 PR 30-NOV-1999: 99WO-US28214.  
 PR 30-NOV-1999: 99WO-US28313.  
 PR 01-DEC-1999: 99WO-US28409.  
 PR 01-DEC-1999: 99WO-US28301.  
 PR 02-DEC-1999: 99WO-US28634.  
 PR 02-DEC-1999: 99WO-US28551.  
 PR 02-DEC-1999: 99WO-US28564.  
 PR 16-DEC-1999: 99WO-US28565.  
 PR 20-DEC-1999: 99WO-US30095.  
 PR 30-DEC-1999: 99WO-US30999.  
 PR 05-JAN-2000: 99WO-US31274.  
 PR 06-JAN-2000: 2000WO-US00219.  
 PR 06-JAN-2000: 2000WO-US00277.  
 PR 11-FEB-2000: 2000WO-US00376.  
 PR 18-FEB-2000: 2000WO-US03565.  
 PR 18-FEB-2000: 2000WO-US04341.  
 PR 22-FEB-2000: 2000WO-US04342.  
 PR 22-FEB-2000: 2000WO-US04414.  
 XX  
 PA (GENE) GENFITECH INC.  
 XX  
 PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W,  
 PI Kadakoff RC, Lu Y, Pan J, Pennica D, Shelton DL, Smith V,  
 PI Stewart TA, Thmas D, Matanabe CK, Wood WI, Yan M;  
 PI  
 XX WPI: 2000-572271/53.  
 DR N-PSDB: AAC58584.  
 XX  
 PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of  
 PT immune related disorders, e.g. systemic lupus erythematosus, rheumatoid  
 PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus.  
 XX  
 PS Claim 33: Fig 12: 309pp: English.  
 XX  
 CC The present invention describes sixty four human PRO proteins which can  
 CC be used in the treatment of immune related diseases. The human PRO  
 CC proteins, anti-PRO antibodies, agonists and antagonists are useful for  
 CC treating and diagnosing immune related disorders. The disorders are  
 CC selected from systemic lupus erythematosus, rheumatoid arthritis,  
 CC osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,  
 CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's  
 CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic  
 CC anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus,  
 CC immune-mediated renal disease, demyelinating diseases of the central  
 CC and peripheral nervous systems, hepatobiliary diseases, inflammatory  
 CC bowel disease, gluten-sensitive enteropathy and Whipple's disease,  
 CC autoimmune or immune-mediated skin diseases, allergic diseases,  
 CC immunological diseases of the lung, and transplantation associated  
 CC diseases including graft rejection and graft-versus-host-disease.  
 CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used  
 CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and  
 CC AAB33414 to AAB33477 represent human PRO polynucleotide and protein  
 CC sequences given in the exemplification of the present invention.  
 XX  
 SQ Sequence 229 AA;  
 Query Match 100.0%; Score 309; DB 21; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-25;

Matches 52: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QHVNCPGGINAMNTITSTYIDNOCGOKNLCNNTGPEMCPENGSCVPPGPG 52  
 126 qhvnpcpgginawntitsyidnqlcgqknlcntgtgpcmcpcngscvppdpg 177

## RESULT 5

AA88570 standard; Protein; 229 AA.

AA88570;

09-AUG-2000 (first entry)

Human PRO240 amino acid sequence.

Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;

PRO317; tumour growth inhibitor; cancer; diagnosis; treatment; human;

cell growth proliferation; serrate precursor; C-serrate-1; ADEPT;

antibody dependent enzyme mediated prodnrg therapy; chromosome 2.

Homo sapiens.

WO200015666-A2.

23-MAR-2000.

08-SEP-1999; 99WO-US20594.

10-SEP-1998; 98US-0099803.

10-SEP-1998; 98WO-US18824.

(GETH ) GENENTECH INC.

Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI, Botstein D;

WPI: 2000-271386/23.

N-PSDB; AAA30036.

New isolated antibodies which bind to specific polypeptides used for

diagnosis and treatment of neoplastic cell growth and proliferation.

Example 4; Fig 8; 200pp; English.

This sequence represents a human PRO240 amino acid sequence. PRO240

shares sequence homology with the D. melanogaster serrate precursor

protein and the Gallus gallus C-serrate-1 protein. The PRO240 gene is

located on chromosome 2. The invention relates to isolated antibodies

which bind to a polypeptide. The "PRO" polypeptides are encoded by genes

which are over expressed in the genome of tumour cells. Vectors and host

cells comprising the nucleic acid encoding the antibodies are used in the

production of the antibodies. The antibodies and nucleic acids encoding

CC them are used for diagnosing a tumour in a mammal. The antibodies are

used for inhibiting the growth of tumour cells and identifying compounds

of a PRO187, PRO533, PRO214, PRO240, PRO211, PRO230, PRO261, PRO246 or

PRO317 polypeptide. The antibody can be used in antibody dependent enzyme

mediated prodnrg therapy (ADEPT) by conjugating the antibody to a

prodnrg-activating enzyme which converts a prodnrg to an anti-cancer

drug. The antibodies can be fluorescently labelled and monitored by light

Db 126 qhvnpcpgginawntitsyidnqlcgqknlcntgtgpcmcpcngscvppdpg 177

## RESULT 6

AA04295 standard; Protein; 229 AA.

AA04295;

24-OCT-2001 (first entry)

Transforming growth factor (TGF) alpha HIII.

Human; TGF alpha HIII; transforming growth factor alpha HIII; cancer;

diagnostic; therapeutic; immune disorder; multiple sclerosis;

systemic lupus erythematosus; human immuno-deficiency virus; HIV;

hyperproliferative disorder; Gaucher's disease; cardiovascular disease;

Scimitar syndrome; Chaga's cardiomyopathy; coronary arteriosclerosis;

angiogenic disorder; corneal graft; neovascularisation; wound healing;

diabetic retinopathy; neurological disorder; Huntington's chorea;

Alzheimer's disease; Parkinson's disease.

Homo sapiens.

Key

Peptide

Protein

Active-site

Region

WO200140251-A1.

07-JUN-2001.

01-DEC-2000; 2000WO-US32745.

02-DEC-1999; 99US-0168387.

(HUMA-) HUMAN GENOME SCI INC.

Wei Y;

WPI: 2001-441480/47.

N-PSDB; AAS08543.

Nucleic acid encoding human transforming growth factor alpha III

(TGFa), useful for preventing, diagnosing and/or treating e.g. Cancer

and Parkinson's disease.

Claim 11; Fig 1; 302pp; English.

The sequence represents the amino acid sequence of human transforming

growth factor (TGF) alpha HIII. TGF alpha HIII nucleic acid and protein

may be used in the prevention, diagnosis and treatment of diseases

associated with inappropriate polypeptide expression, for example immune

disorders (e.g. multiple sclerosis, systemic lupus erythematosus and

human immuno-deficiency virus (HIV) infections), hyperproliferative

disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases

(e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary

arteriosclerosis), angiogenic disorders (e.g. corneal graft

neovascularisation, wound healing, Huntington's chorea;

(e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),

infectious diseases and/or for promoting wound healing, regeneration

and/or chemotaxis (full details given in specification). Additionally,

the nucleic acid may be used to produce the secreted polypeptides, by

inserting the nucleic acids into a host cell and culturing the cell to

express the protein. It may also be used as a DNA probe in diagnostic

assays to detect and quantitate the presence of similar nucleic acid

sequences in samples, and therefore which patients may be in need of

restorative therapy. The polypeptides may also be used as antigens in the

CC production of antibodies against TGF alpha HIII and in assays to identify  
 CC modulators of TGF alpha HIII. The anti-TGF alpha HIII antibodies may also  
 CC be used as diagnostic agents for detecting the presence of TGF alpha HIII  
 CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)).  
 XX

SO Sequence 229 AA:

Query Match 100.0%; Score 309; DB 22; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-25;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 OHVNCPEGINAMNTITSYDNOICGOKNLCNNTGDPKPCPNKSCVPDGP 52  
 126 qhncpgglnawntitsyldngicggknlcnntgdpemcpcngscvpdpgp 177

RESULT 7  
 AAB20112

ID AAB20112 standard; Protein; 229 AA.

XX AAB20112;

XX 30-APR-2001 (first entry)

XX Human immunostimulant PRO240.

DE PRO240; UNQ214; human; immune disease; autoimmune disease;  
 KW antirheumatic; antiarthritic; antinflammatory; antianemic;  
 KW immunosuppressive; antithyroid; antidiabetic; neuroprotective;  
 KW hepatotropic; virucide; dermatological; antiproliferative;  
 KW antiallergic; antiallergic; immunostimulant; serrate; lung cancer.  
 XX Homo sapiens.

OS Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..30

FT Protein 31..229

FT Domain 198..213

FT Region 181..193

FT Modified-site 44..48 /note= "epidermal growth factor domain cysteine pattern signature"

FT Modified-site 79..83 /note= "Asn is N-glycosylated"

FT Modified-site 157..161 /note= "Asn is N-glycosylated"

FT Modified-site 168..172 /note= "Asn is N-glycosylated"

FT Modified-site 7..13 /note= "Asn is N-glycosylated"

FT Modified-site 24..30 /note= "N-myristoylation site"

FT Modified-site 40..46 /note= "N-myristoylation site"

FT Modified-site 70..76 /note= "N-myristoylation site"

FT Modified-site 132..138 /note= "N-myristoylation site"

FT Modified-site 177..183 /note= "N-myristoylation site"

FT Modified-site 205..211 /note= "N-myristoylation site"

XX Modified-site /note= "N-myristoylation site"

PN W0200105972-A1.

XX 25-JAN-2001.

XX 15-MAR-2000; 2000WO-US06884.

XX 20-JUL-1999; 99US-0144758.

XX (SETH ) GENENTECH INC.

PI Ashkenazi AJ, Baker KP, Fong S, Goddard A, Godowski PJ, Gurney AL,

PI Hillan KJ, Mark MR, Marsters SA, Pitti RM, Tumas D, Watanabe CK;

XX Wood WI;

XX WPI; 2001-103149/11.

XX N-PSDB; AAF30054.

PS Claim 20; Fig 10; 127pp; English.

CC The present sequence is that of novel human immunomodulator PRO240  
 CC (UNQ214), as deduced from cDNA (see AAF30054) isolated from a  
 CC foetal liver library. PRO240 (25 kDa, pI 7.83) shows sequence  
 CC homology to chicken C-serrate-1 and Drosophila serrate precursor  
 CC protein. Expression was observed in lung cancer, 8 squamous  
 CC carcinomas and in 6/8 adenocarcinomas, in situ and infiltrating  
 CC components. The invention provides polynucleotides (see AAF30050-62)  
 CC encoding novel human PRO proteins (see AAB20108-20) including  
 CC PRO240. Claimed compositions comprising these proteins or their  
 CC agonists are useful for increasing infiltration of inflammatory  
 CC cells into a tissue of a mammal, stimulating or enhancing an immune  
 CC response, or increasing the proliferation of T-lymphocytes in a  
 CC mammal in response to an antigen. Claimed compositions comprising  
 CC a PRO polypeptide or its antagonist have the opposite effect. A  
 CC claimed method for treating an immune related disorder, such as a T  
 CC cell disorder, involves administering a PRO polypeptide, an agonist  
 CC antibody or an antagonist antibody. The disorder is selected from  
 CC systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,  
 CC juvenile chronic arthritis, spondyloarthritis, systemic sclerosis,  
 CC vasculitis, sarcoidosis, autoimmune haemolytic anaemia, autoimmune  
 CC thrombocytopenia, thyroiditis, diabetes mellitus, immune-mediated  
 CC renal disease, demyelinated diseases (such as multiple sclerosis),  
 CC granulomatous hepatitis, sclerosing cholangitis, inflammatory bowel  
 CC disease (ulcerative colitis and Crohn's disease), gluten sensitive  
 CC enteropathy, Whipple's disease, (auto)immune-mediated skin diseases  
 CC (such as bullous skin disease, erythema multiforme and psoriasis),  
 CC allergic diseases (such as asthma, allergic rhinitis, atopic  
 CC dermatitis, food hypersensitivity and urticaria), immunologic  
 CC diseases of the lung and transplantation associated diseases (such  
 CC as graft rejection and graft-versus-host disease) (all claimed).  
 CC Claimed methods of diagnosing these disorders comprise detecting  
 CC the level of expression of the PRO gene. Also claimed are a method  
 CC of identifying a compound capable of inhibiting the expression or  
 CC activity of the PRO polypeptide, vectors, host cells, antibodies,  
 CC and a method of stimulating an immune response in a mammal using  
 CC PRO240.

XX Sequence 229 AA:

Query Match 100.0%; Score 309; DB 22; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-25;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 OHVNCPEGINAMNTITSYDNOICGOKNLCNNTGDPKPCPNKSCVPDGP 52  
 126 qhncpgglnawntitsyldngicggknlcnntgdpemcpcngscvpdpgp 177

RESULT 8

XX AAB68595

XX 15-MAR-2000; 2000WO-US06884.

AC AAB68595;  
 XX 27-APR-2001 (first entry)  
 XX PRO240.  
 DE Cytostatic; PRO protein; tumour; cancer.  
 KW Homo sapiens.  
 XX MO200105836-A1.  
 XX 25-JAN-2001.  
 PD 20-DEC-1999; 99WO-US30999.  
 XX 20-DEC-1999; 99US-0144758.  
 XX 26-JUL-1999; 99US-0145698.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.  
 PR 05-SEP-1999; 99WO-US23089.  
 PR 05-OCT-1999; 99WO-US28214.  
 PR 29-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28564.  
 XX (GENTH ) GENENTECH INC.  
 XX Botstein D, Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WI;  
 PI WPI: 2001-091968/10.  
 DR N-PSDB; AAF60356.  
 XX New antibody that binds to a PRO polypeptide, e.g. PRO187 and PRO533,  
 PT useful for diagnosing and treating cancers -  
 XX Claim 61; Fig 8; 196pp; English.  
 XX The present invention relates to PRO proteins and coding sequences. The  
 CC present sequence is one such PRO protein. It was found that the PRO genes  
 CC are amplified in the genome of tumour cells. The gene amplification is  
 CC expected to be associated with the overexpression of the gene product and  
 CC contributes to tumorigenesis. Therefore, antagonists of PRO proteins are  
 CC useful for the treatment of benign or malignant tumours, leukaemias,  
 CC lymphoid malignancies and other disorders such as neuronal, glial,  
 CC astrocytic, hypothalamic, glandular, epithelial, inflammatory and  
 CC immunologic disorders.  
 XX Sequence 229 AA;  
 SQ  
 OY 1 QHVNCPGGINAMNTTTSYIDNOCGOKNLNNTGDPENCGSCVPDPG 52  
 DB 126 qhvnepgginawntitsyidngicgqknlcntgdpemcpgscvpdp 177  
 RESULT 9  
 AAB50951  
 ID AAB50951 standard; Protein; 229 AA.  
 XX AAB50951;  
 AC 21-MAR-2001 (first entry)  
 DE Human PRO240 protein.  
 XX Human; PRO; cytotostatic; nootropic; neuroprotective; respiratory general;  
 KW antiinflammatory; antiangiogenic; immunosuppressive; immunostimulant;  
 KW PRO agonist; cancer; inflammatory disorder; immunological disorder.

XX Homo sapiens.  
 OS MO200073348-A2.  
 PN 07-DEC-2000.  
 PD 30-MAY-2000; 2000WO-US14941.  
 PF 02-JUN-1999; 99WO-US12252.  
 PR 22-JUN-1999; 99US-0140650.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 01-DEC-1999; 99WO-US28634.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 20-DEC-1999; 99WO-US30999.  
 PR 06-JAN-2000; 2000WO-US00376.  
 PR 11-FEB-2000; 2000WO-US03365.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 02-MAR-2000; 2000US-0187202.  
 PR 03-MAR-2000; 2000WO-US06319.  
 PR 10-MAR-2000; 2000WO-US06884.  
 PR 15-MAR-2000; 2000WO-US08439.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 17-MAY-2000; 2000WO-US13705.  
 XX (GENTH ) GENENTECH INC.  
 XX Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W, Kabakoff RC;  
 PI Shelton DL, Smith V, Watanabe CX, Wood WI;  
 DR WPI: 2001-016509/02.  
 DR N-PSDB; AAC91553.  
 XX Twenty eight nucleic acids encoding PRO polypeptides which are useful  
 PT for treating various tumors, e.g. breast cancer, and other  
 PT inflammatory, angiogenic and immunological disorders -  
 XX Claim 31; Fig 2; 186pp; English.  
 XX The present sequence is one of twenty eight novel PRO polypeptides. The  
 CC PRO polypeptides and their agonists, including antibodies, peptides, and  
 CC small molecule agonists, may be used to treat various tumours, e.g.,  
 CC cancers such as breast cancer, prostate cancer, lung cancer, bladder cancer,  
 CC cancer, uterine cancer, ovarian cancer, renal cancer, colorectal  
 CC central nervous system cancer, melanoma, or leukaemia. They are also  
 CC useful for treating other disorders such as neuronal, glial, astrocytic,  
 CC hypothalamic and other glandular, macrophagal, epithelial, stromal and  
 CC blastocoeleic disorders, and inflammatory, angiogenic and immunological  
 CC disorders.  
 XX Sequence 229 AA;  
 SQ  
 OY 1 QHVNCPGGINAMNTTTSYIDNOCGOKNLNNTGDPENCGSCVPDPG 52  
 DB 126 qhvnepgginawntitsyidngicgqknlcntgdpemcpgscvpdp 177  
 RESULT 10  
 AAB48068  
 ID AAB48068 standard; protein; 342 AA.  
 XX



AA048068;  
 19-MAR-2001 (first entry)  
 Human extracellular signaling molecule (EXCS) (ID 2207183CD1).  
 Extracellular signaling molecule; EXCS; anti-inflammatory; human; immunosuppressive; cytostatic; neuroprotective; gastrointestinal; antiviral; antibacterial; anti-HIV; human immunodeficiency virus; antileishmanial; cerebroprotective; neurotropic; antitumor; antifungal; anticonvulsant; tranquilizer; neuroleptic; vasotropic; gynecological; keratolytic; protozoacide; gene therapy.  
 Homo sapiens.  
 WO200070049-A2.  
 23-NOV-2000.  
 19-MAY-2000; 2000WO-US13975.  
 19-MAY-1999; 99US-0134949.  
 15-JUL-1999; 99US-0144270.  
 30-JUL-1999; 99US-0146700.  
 04-OCT-1999; 99US-0157508.  
 (INCYTE) INCYTE GENOMICS INC.  
 Tang YT, Yue H, Lal P, Burford N, Bandman O, Baughn MR; Azimzai Y, Lu DM, Paterson C;  
 WPI: 2001-025021/03.  
 N-PDB: AAC84304.  
 New human extracellular signaling nucleic acids and polypeptides useful for diagnosing, treating and preventing infections and gastrointestinal, neurological, reproductive, and autoimmune/inflammatory disorders -  
 Claim 1; Page 89-90; 114pp; English.  
 The invention provides human extracellular signaling molecules (EXCS) and polynucleotides which identify and encode EXCS. EXCS can be expressed by standard recombinant methodology. The amino acid and nucleic acid sequences of EXCS are useful for diagnosing, treating and preventing infections and gastrointestinal (peptic ulcer, dysphagia, pancreatitis), neurological (e.g. epilepsy, ischemic cerebrovascular disease, stroke), reproductive (infertility, ovulatory defects, endometriosis), autoimmune/inflammatory (actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease), and cell proliferative disorders including cancers (of the breast, adrenal gland, bone). They may also be used to treat fatal familial insomnia, nutritional and metabolic diseases of the nervous system, myopathies, mental disorders (anxiety, schizophrenia, mood), as well as infections caused by parasites (malaria, leishmania, trypanosoma), viral (adenovirus, coronavirus, flavivirus), bacterial (e.g. pneumococcus, staphylococcus, bacillus), and fungal (aspergillus, blastomycosis, dermatophytes) agents. The nucleic acids, polypeptides, antagonists, agonists, pharmaceutical compositions, and antibodies may also be used for treating or preventing disorders associated with increased or decreased expression or activity of EXCS. EXCS polynucleotides may also be used to detect and quantify gene expression in biopsied tissues in which expression of EXCS may be correlated with the disease. To determine presence or excess expression of EXCS, to monitor regulation of EXCS levels during therapeutic intervention, to detect the presence of EXCS associated disorders, as targets in microarray, to generate hybridization probes, and to detect differences in gene sequences among normal, carrier or affected individuals. Antibodies may also be used in diagnosing disorders, in monitoring patients being treated with EXCS agonists, antagonists or inhibitors. Sequences AA048057-B48082 represent the EXCS of the invention.

Query Match 100.0%; Score 309; DB 22; Length 342;  
 Best Local Similarity 100.0%; Pred No. 2.2e-25;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QHWNCPGGINAMWNTTYSIDNQCQOQKNCNNTGPEMCPENGSCVPPDGP 52  
 Db 239 ghvncpgjlnawntltstydngicqgknlcmntgdpemcpgnscvppdgp 290  
 RESULT 11  
 AA004296 standard; Protein; 52 AA.  
 AA004296;  
 24-OCT-2001 (first entry)  
 Transforming growth factor (TGF) alpha HI.  
 Human; TGF alpha HI; transforming growth factor alpha HIII; cancer; diagnostic; therapeutic; immune disorder; multiple sclerosis; HIV; systemic lupus erythematosus; human immuno-deficiency virus; HIV; hyperproliferative disorder; Gaucher's disease; cardiovascular disease; Scintlar syndrome; Chaga's cardiomyopathy; coronary arteriosclerosis; angiogenic disorder; corneal graft; neovascularisation; wound healing; diabetic retinopathy; neurological disorder; Huntington's chorea; Alzheimer's disease; Parkinson's disease.  
 Homo sapiens.  
 WO200140251-A1.  
 07-JUN-2001.  
 01-DEC-2000; 2000WO-US32745.  
 02-DEC-1999; 99US-0168387.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Wei Y;  
 WPI: 2001-441480/47.  
 Nucleic acid encoding human transforming growth factor alpha III (TGF $\alpha$ ), useful for preventing, diagnosing and/or treating e.g. Cancer and Parkinson's disease -  
 Disclosure; Fig 2; 302pp; English.  
 The sequence represents the amino acid sequence of human transforming growth factor (TGF) alpha HI. TGF alpha HI nucleic acid and protein may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression, for example immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus and human immuno-deficiency virus (HIV) infections), hyperproliferative disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases (e.g. Scintlar syndrome, Chaga's cardiomyopathy), hyperproliferative arteriosclerosis), angiogenic disorders (e.g. coronary artery disease), neovascularisation and diabetic disorders (e.g. corneal graft, neovascularisation and diabetic retinopathy), neurological disorders (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease), and/or chemotaxis (full details given in specification). Additionally, the nucleic acid may be used to produce the secreted polypeptides, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. It may also be used as a DNA probe in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples, and therefore which patients may be in need of restorative therapy. The polypeptides may also be used as antigens in the production of antibodies against TGF alpha HIII and in assays to identify modulators of TGF alpha HIII. The anti-TGF alpha HIII antibodies may also

CC be used as diagnostic agents for detecting the presence of TGF alpha HIII  
 CC in samples (e.g. by enzyme linked immunosorbant assay (ELISA)).  
 CC

Sequence 52 AA;

Query Match 52.8%; Score 163; DB 22; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 GOKNLGNTGDPMPGSCVPDPG 52  
 ||||||||||||||||||  
 Db 1 gqknlcnntgdpmpgscvpdp 27

RESULT 12

AAG00157  
 ID AAG00157 standard; Protein; 147 AA.

AC AAG00157;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 4238.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.

OS Homo sapiens.

PN EPI033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR N-PSDB; AANC00163.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 13; SEQ ID 4238; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30,  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.

Sequence 147 AA;

Query Match 41.4%; Score 128; DB 21; Length 147;

Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 OHVNCPGINAWNTITSYIDNQ 22

Db 126 gvhncpginawntitsyidnq 147  
 ||||||||||||||||||

RESULT 13

ID AAY27073 standard; Protein; 296 AA.

AC AAY27073;

DT 18-OCT-1999 (first entry)

DE L. bimastus plasmin mature protein sequence.

XX Lumbricus bimastus; earthworm; plasmin gene.

OS Lumbricus bimastus.

PH Key Location/Qualifiers

FT Misc-difference 4 /note= "encoded by GC"

FT Misc-difference 12 /note= "encoded by TGC"

FT Misc-difference 243 /note= "encoded by TAG"

FT Misc-difference 268 /note= "encoded by TAA"

FT Misc-difference 274 /note= "encoded by TAA"

FT Misc-difference 279 /note= "encoded by TAA"

FT Misc-difference 280 /note= "encoded by TAA"

FT Misc-difference 292 /note= "encoded by TAA"

FT Misc-difference 296 /note= "encoded by TAG"

CN108770-A.

PD 24-FEB-1999.

XX 11-JUN-1998; 98CN-0102257.

PR 11-JUN-1998; 98CN-0102257.

PA (VIR-) VIROLOGY RES INST CHINA PREVENTIVE MEDIC.

PI Fu S, Liang G, Meng X;

DR WPI: 1999-313741/27.

DR N-PSDB; AAX89869.

XX Lumbrical fibrinolysin gene nucleotide series, and method for clone  
 PT of same - is composed of 88 nucleotides in which position 1-726  
 PT nucleotide is gene matured peptide sequence, position 727-729 nucleotide  
 PT nucleotide is terminal codon TAG

PS Claim 3; Page 1; 9pp; Chinese.

XX The invention provides a Lumbricus bimastus (a kind of earthworm) plasmin  
 CC gene. The gene is composed of 88 nucleotides in which position 1-726  
 CC nucleotide is gene matured peptide sequence, position 727-729 nucleotide  
 CC is terminal codon TAG. The expressed protein from the gene possesses  
 CC tremendous medical prospect. The present sequence represents the  
 CC L. bimastus plasmin gene mature protein sequence.

Sequence 296 AA;

Query Match 24.9%; Score 77; DB 20; Length 296;

Best Local Similarity 30.3%; Pred. No. 1.1;  
 Matches 20; Conservative 5; Mismatches 13; Indels 28; Gaps 4;

OY 7 GGIAA---NTTTSYIDNQC-----OGKRLCN-NTGDEPCPEN-----CS 45  
 DB 166 gganitw-----dnhicvqdpagntgacngdsqplncpdggttrvvtswvssqjlt 218  
 OY 46 CVPDGP 51  
 DB 219 clpdy 224

## RESULT 14

AA067913  
 ID AAR67913 standard; Protein; 2019 AA.

AC AAR67913;

DT 05-AUG-1995 (first entry)

DE Cardiac sodium channel protein.

KW Sodium channel protein; therapeutic; diagnostic; prognostic;  
 antiarrhythmic; cardiant; cardioglycoside.

OS Rattus rattus.

PN US5380836-A.

PD 10-JAN-1995.

PF 13-FEB-1989; 89US-0331330.

PR 13-FEB-1989; 89US-0331330.

PR 30-SEP-1991; 91US-0768107.

PA (ARCH-) ARCH DEV CORP.

PI Rogart RB;

DR WPI; 1995-060381/08.

DR P-PSDB; AA081328.

PT Purified DNA's encoding rat and human cardiac sodium channel  
 protein - useful for recombinant expression to produce sodium

PT channel proteins.

PS Disclosure; Fig 2; 39pp; English.

CC The rat cardiac channel protein has various therapeutic, diagnostic  
 and prognostic uses. It may also be used to develop more effective  
 antiarrhythmic, cardiant and cardioglycoside drugs. In Figure 2,  
 CC the sequence is compared to the deduced amino acid sequence of rat  
 CC brain II cDNA.

SO Sequence 2019 AA;

Query Match 21.7%; Score 67; DB 16; Length 2019;  
 Best Local Similarity 29.6%; Pred. No. 99;

Matches 16; Conservative 8; Mismatches 20; Indels 10; Gaps 2;

OY 7 GGIAA---NTTTSYIDNQC-----LCNNTGDEPCPENSGCVPDGP 50  
 DB 293 gsvaadglvwnsladvlyndpanyllkngttdvllcgnsdadtcegyrclxag 346

## RESULT 15

AA06584

ID AAR06584 standard; protein; 2020 AA.

AC AAR06584;

DT 10-JAN-1991 (first entry)

XX

DE Cardiac sodium channel.

KW Rat; arrhythmia.

OS Rattus rattus.

PN W09009391-A.

PD 23-AUG-1990.

PF 09-FEB-1990; 90MO-US00768.

PR 13-FEB-1989; 89US-0310330.

PA (ARCH-) ARCH DEV CORP.

PI Rogart RB;

DR WPI; 1990-275095/36.

DR N-PSDB; AA05831.

PT New rat cardiac sodium channel proteins - and associated DNA  
 sequences, polypeptides and peptides associated with  
 PT proteins, useful as antiarrhythmic and cardiotoxic drugs.

PS Disclosure; Fig 2; 65pp; English.

CC The sequence deduced from cDNA derived from 3 overlapping clones,  
 CC PRH3-1, PRH4-23, and PRH4-31. (Deposited as ATCC 67885, 67886,  
 CC and 67887 resp.) The clones were isolated from a cDNA library in  
 CC the lambda zap vector prep. from mRNA obtd. from newborn rat  
 CC hearts using rat brain II cDNA probe. The protein has diagnostic,  
 CC therapeutic, and prognostic applications.

SO Sequence 2020 AA;

Query Match 21.7%; Score 67; DB 11; Length 2020;  
 Best Local Similarity 29.6%; Pred. No. 1e+02;

Matches 16; Conservative 8; Mismatches 20; Indels 10; Gaps 2;

OY 7 GGIAA---NTTTSYIDNQC-----LCNNTGDEPCPENSGCVPDGP 50  
 DB 293 gsvca\$lvwnsladvlyndpanyllkngttdvllcgnsdaa'cegyrclxag 346

Search complete: March 28, 2002, 08:58:43  
 Job time: 5170

Thu Mar 28 09:21:37 2002

us-09-726-348-2\_copy\_126\_177\_1.rag

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:15:53 ; Search time 59.64 Seconds

(without alignments)  
64.584 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_126\_177

Sequence: 1 QHVNCPGGINAMNTTTSYID.....NTGDPENCPCPENGSCVDPDGP 52

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 237521

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	11.5	10	AA11404	C. maltosa cytochr
2	11.5	15	AA11403	C. maltosa cytochr
3	11.5	25	AAW60083	M. vaccae antigen
4	11.5	25	AAV14821	N-terminal fragmen
5	9.6	6	AA874817	Cyclopropylbenzind
6	9.6	7	AA870878	Vasopressin (ADH)
7	9.6	7	AA870878	Vasopressin (ADH)
8	9.6	7	AA870878	Vasopressin (ADH)
9	9.6	7	AA870878	Vasopressin (ADH)
10	9.6	9	AA870878	Consensus peptide,
11	9.6	10	AAV06014	Clan B-cell epit
				Human Cancer anti

12	5	9.6	10	22	AA87571	Human complementar
13	5	9.6	10	22	AA87572	Human complementar
14	5	9.6	10	22	AA87784	Saccharomyces cere
15	5	9.6	10	22	AA87785	Saccharomyces cere
16	5	9.6	10	22	AA87786	Saccharomyces cere
17	5	9.6	10	22	AA87787	Saccharomyces cere
18	5	9.6	10	22	AA87788	Saccharomyces cere
19	5	9.6	10	22	AA87789	Saccharomyces cere
20	5	9.6	10	22	AA87790	Saccharomyces cere
21	5	9.6	10	22	AA87791	Saccharomyces cere
22	5	9.6	10	22	AA87792	Saccharomyces cere
23	5	9.6	10	22	AA87793	Saccharomyces cere
24	5	9.6	10	22	AA87794	Saccharomyces cere
25	5	9.6	10	22	AA87795	Saccharomyces cere
26	5	9.6	10	22	AA87796	Saccharomyces cere
27	5	9.6	10	22	AA87797	Saccharomyces cere
28	5	9.6	10	22	AA87798	Saccharomyces cere
29	5	9.6	10	22	AA87799	Saccharomyces cere
30	5	9.6	10	22	AA87800	Saccharomyces cere
31	5	9.6	10	22	AA87801	Saccharomyces cere
32	5	9.6	10	22	AA87802	Saccharomyces cere
33	5	9.6	10	22	AA87803	Saccharomyces cere
34	5	9.6	10	22	AA87804	Saccharomyces cere
35	5	9.6	10	22	AA87805	Saccharomyces cere
36	5	9.6	10	22	AA87806	Saccharomyces cere
37	5	9.6	10	22	AA87807	Saccharomyces cere
38	5	9.6	10	22	AA87808	Saccharomyces cere
39	5	9.6	10	22	AA87809	Saccharomyces cere
40	5	9.6	10	22	AA87810	Saccharomyces cere
41	5	9.6	10	22	AA87811	Saccharomyces cere
42	5	9.6	10	22	AA87812	Saccharomyces cere
43	5	9.6	10	22	AA87813	Saccharomyces cere
44	5	9.6	10	22	AA87814	Saccharomyces cere
45	5	9.6	10	22	AA87815	Saccharomyces cere

## ALIGNMENTS

RESULT 1	AA11404	standard: peptide; 10 AA.
AA11404:		
22-FEB-2001 (first entry)		
C. maltosa cytochrome b5 peptide fragment T54.		
Cytochrome b5 alkane metabolism; oxidation; long-chain alkyl compound;		
Long-chain alkyl compound.		
Candida maltosa		
W0200065061-42.		
02-NOV-2001		
18-APR-2000; 2000MO-DE01246.		
24-APR-1999; 99DE-1018763.		
(DELB-) DELBROECK CENT MOLEKULARE MEDIZIN MAX.		
Schunck W, Chernogolov A;		
WPI; 2000-679674/66.		
Nucleic acid sequences from alkane metabolizing Candida yeast, encoding		
cytochrome b5 and used for the oxidation of long chain alkyl compounds		
and for the production of long chain dicarboxylic acids -		
Example 1.2; Page 6; 27pp; German.		

XX This invention describes novel nucleic acid sequences from alkane  
 CC metabolizing Candida yeasts, encoding cytochrome b5-polypeptide, its  
 CC fragments, variants and mutations. The nucleic acids and polypeptides  
 CC are used for the oxidation of long-chain alkyl compounds with at least  
 CC 10C, and for the production of long-chain dicarboxylic acids by oxidizing  
 CC n-alkanes and fatty acids with at least 10C.  
 CC  
 XX  
 SO Sequence 10 AA:

Query Match 11.5%; Score 6; DB 21; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 ITSYPD 20  
 |||||  
 DB 4 Itsyid 9

## RESULT 2

AAB11403 standard; peptide: 15 AA.

XX AAB11403;

DT 22-FEB-2001 (first entry)

DE C. maltosa cytochrome b5 peptide fragment T53.

KW Cytochrome b5; alkane metabolism; oxidation; long-chain alkyl compound;  
 KW long-chain dicarboxylic acid.

OS Candida maltosa.

PN WO200065061-A2.

PD 02-NOV-2000.

PF 18-APR-2000; 2000WO-DE01246.

PR 24-APR-1999; 99DE-1018763.

PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

PI Schunck W, Chernogolov A;

DR WPI; 2000-679674/66.

XX Nucleic acid sequences from alkane metabolizing Candida yeast, encoding  
 PT cytochrome b5 and used for the oxidation of long chain alkyl compounds  
 PT and for the production of long chain dicarboxylic acids -  
 XX

PS Example 1.2; Page 6; 27pp; German.

XX This invention describes novel nucleic acid sequences from alkane  
 CC metabolizing Candida yeasts, encoding cytochrome b5-polypeptide, its  
 CC fragments, variants and mutations. The nucleic acids and polypeptides  
 CC are used for the oxidation of long-chain alkyl compounds with at least  
 CC 10C, and for the production of long-chain dicarboxylic acids by oxidizing  
 CC n-alkanes and fatty acids with at least 10C.  
 CC  
 XX

SO Sequence 15 AA:

Query Match 11.5%; Score 6; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 8.2;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 ITSYPD 20  
 |||||  
 DB 4 Itsyid 9

## RESULT 3

AAM60083 standard; protein: 25 AA.

XX AAM60083;

DT 25-AUG-1998 (first entry)

DE M. vaccae antigen Gvc-1 N-terminal sequence.

KW Mycobacterium vaccae; antigen; therapy; prevention; cytokine production;  
 KW avium; M. tuberculosis; immune response enhancer; cell proliferation;  
 KW mycobacteria infection; vaccine; cancer.

OS Mycobacterium vaccae.

XX Key Location/Qualifiers  
 FT Misc-difference 7 /note="unknown"

PN WO9808542-A2.

PD 05-MAR-1998.

PF 28-AUG-1997; 97WO-NZ00105.

PR 12-JUN-1997; 97US-0873970.

PR 29-AUG-1996; 96US-0705347.

PA (GENE-) GENESIS RES & DEV CORP.

PI Hiyama J, Prestidge RL, Scott LM, Skinner MA, Tan P;  
 PI Visser E;

DR WPI; 1998-216926/19.

XX Mycobacterium vaccae polypeptides - used to develop products for use  
 PT in detection, therapy and prevention of mycobacteria infections or  
 PT as immune response enhancers  
 XX

PS Claim 2; Page 61; 153pp; English.

XX This represents the Mycobacterium vaccae antigen Gvc-1 N-terminal  
 CC sequence. The invention provides M. vaccae polypeptides that comprise  
 CC an immunogenic portion of a soluble M. vaccae antigen, or a variant,  
 CC where the antigen induces an immune response in patients previously  
 CC exposed to a mycobacterium. Such M. vaccae polypeptides can be used in  
 CC methods for enhancing non-specific immune response. The methods and  
 CC products can be used for the detection, treatment and prevention of  
 CC infectious diseases caused by mycobacteria such as M. vaccae, M. avium  
 CC or M. tuberculosis. The products also have the ability to induce cell  
 CC proliferation and cytokine production (e.g. Interferon-gamma and  
 CC interleukin-12 production) in T cells, NK cells, B cells, or macrophages.  
 CC They can be used for enhancing immune responses for use in vaccines or  
 CC immunotherapy of infectious diseases and cancers.  
 CC  
 XX

SO Sequence 25 AA:

Query Match 11.5%; Score 6; DB 19; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VPDGPG 52  
 |||||  
 DB 14 vpdgpg 19

## RESULT 4

AAV14821 standard; peptide: 25 AA.

XX

AC AAY14821;  
 XX  
 DT 25-OCT-1999 (first entry)  
 XX  
 DE N-terminal fragment of M. vaccae antigen GVC-1.  
 XX  
 KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;  
 KW dendritic cell maturation; infectious disease; immune disorder; cancer;  
 KW respiratory system; mycobacterial infection; allergy; tuberculosis;  
 KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;  
 KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;  
 KW squamous cell carcinoma; melanoma.  
 XX  
 OS Mycobacterium vaccae.  
 XX  
 PN WO932634-A2.  
 XX  
 PD 01-JUL-1999.  
 XX  
 PE 23-DEC-1998; 98WO-NZ00189.  
 XX  
 PR 04-DEC-1998; 98US-0205426.  
 PR 23-DEC-1997; 97US-0966624.  
 PR 23-DEC-1997; 97US-0997080.  
 PR 23-DEC-1997; 97US-0997362.  
 PR 11-JUN-1998; 98US-0095855.  
 PR 17-SEP-1998; 98US-0156181.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 XX  
 PI Prestidge RL, Skinner MA, Tan P, Visser ES, Watson J;  
 DR WPI: 1999-430163/36.  
 XX  
 PT Enhancing immune response to an antigen  
 PT  
 PS Example 10; Page 145; 243pp; English.  
 XX  
 CC The invention provides heat-killed Mycobacterium vaccae, or recombinant  
 CC M. vaccae proteins. The M. vaccae proteins may be employed to activate  
 CC T cells and natural killer cells, to stimulate the production of  
 CC cytokines, to enhance the expression of co-stimulatory molecules on  
 CC dendritic cells and monocytes, and to enhance dendritic cell maturation  
 CC and function. The proteins can be expressed by standard recombinant  
 CC methodology. Pharmaceutical compositions comprising the proteins or  
 CC nucleic acid sequences encoding the proteins can be used for the  
 CC treatment, prevention, and detection of disorders including infectious  
 CC diseases, immune disorders and cancer. In particular, the compounds and  
 CC methods are used for treatment of diseases of the respiratory system,  
 CC such as mycobacterial infections, asthma, allergies, tuberculosis,  
 CC leprosy, sarcoidosis and lung cancers, and disorders of the skin such as  
 CC psoriasis, atopic dermatitis, eczema, allergic contact dermatitis,  
 CC alopecia areata, and skin cancers such as basal carcinoma, squamous cell  
 CC carcinoma and melanoma.  
 CC  
 SQ Sequence 25 AA:  
 QY 47 VPDPG 52  
 DB 14 VPDPG 19  
 11.5%; Score 6; DB 20; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 26-JUN-2001 (first entry)  
 XX  
 DE Cyclopropylbenzindole (CBI) labelled peptide #4.  
 XX  
 KW Cyclopropylindole; cytotoxic; antibiotic; antiparasitic; antiviral;  
 KW therapeutic; benzindole; pyrrolindole.  
 XX  
 OS Synthetic.  
 XX  
 FT Key  
 FT Modified-site 1 Location/Qualifiers  
 FT /note="N-terminally labelled with cyclopropylbenzindole (CBI)"  
 XX  
 PN WO200116104-A1.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PE 24-AUG-2000; 2000WO-GB03291.  
 XX  
 PR 27-AUG-1999; 99GB-0020427.  
 PR 08-MAR-2000; 2000GB-0005576.  
 XX  
 PA (SPIR-) SPIROGEN LTD.  
 XX  
 PI Thurston DE, Howard PW;  
 DR WPI: 2001-307934/32.  
 XX  
 PT Benzo- or pyrrolo-indole derivatives are useful in combinatorial  
 PT methods for discovering cytotoxic, antibiotic, antiparasitic and  
 PT antiviral agents  
 PT  
 PS Example 3; Page 44; 69pp; English.  
 XX  
 CC The present invention describes benzo- or pyrrolo-indole derivatives (I).  
 CC (I) can be used in combinatorial synthesis where it is joined to a  
 CC solid support by a chain comprising at least 1 combinatorial unit. (I)  
 CC can be used in the manufacture of cytotoxic, antibiotic, antiparasitic  
 CC and antiviral therapeutic compositions. The present sequence represents  
 CC a cyclopropylbenzindole (CBI) labelled peptide which is used in an  
 CC example from the present invention.  
 CC  
 SQ Sequence 6 AA:  
 QY 25 OQGN 25  
 DB 1 99qkn =  
 9.6%; Score 5; DB 22; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 4; 3e+05;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6  
 AAB90878  
 ID AAB90878 standard; Peptide: 7 AA.  
 AC AAB90878;  
 XX  
 DT 22-JUN-2001 (first entry)  
 DE Vasopressin (ADH) related peptide SEQ ID NO:50.  
 DE  
 KW Protection; endogenous therapeutic peptide; peptidase; conjugation;  
 KW blood component; modification; succinimidy; maleimido group; amino;  
 KW hydroxy; thiol; hormone; growth factor; neurotransmitter.  
 OS Homo sapiens.  
 OS Synthetic.  
 OS  
 XX  
 PN WO200069900-A2.

23-NOV-2000 .  
17-MAY-2000; 2000OWO-US13576 .  
17-MAY-1999; 99US-0134406 .  
10-SEP-1999; 99US-0153406 .  
15-OCT-1999; 99US-0159783 .  
(CONJ-) CONJUCHEM INC .  
Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudau K;  
WPI; 2001-112059/12 .  
Modifying and attaching therapeutic peptides to albumin prevents  
peptidase degradation, useful for increasing length of in vivo activity  
-  
Disclosure; Page 205; 733pp; English .  
The present invention describes a modified therapeutic peptide (I)  
comprising a therapeutically active amino acid region (III) and a  
reactive group (II) (e.g. succinimidy) and maleimido groups) attached to  
a less therapeutically active amino acid region (IV), which covalently  
bonds with amino/hydroxyl/thiol groups on blood components to form a  
peptide stabilised therapeutic peptide composed of 3-50 amino acids .  
(I) are useful for modifying therapeutic peptides e.g. hormones, growth  
factors and neurotransmitters, to protect them from peptidase activity  
in vivo for the treatment of various disorders . Endogenous therapeutic  
peptides are not suitable as drug candidates as they require frequent  
administration due to rapid degradation by peptidases in the body .  
Modifying and attaching therapeutic peptides to albumin prevents or  
reduces the action of peptidases to increase length of activity (half  
life) and specificity as bonding to large molecules decreases  
intracellular uptake and interference with physiological processes .  
AAB90829 to AAB92441 represent peptides which can be used in the  
exemplification of the present invention .

Query Match	9.6%	Score 5;	DB 22;	Length 7;
Best local Similarity	100.0%	Pred. No. 4.3e+05;		
Matches	5;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;
OY	3	VNCPG 7		
Db	3	vnpcpg 7		
RESULT	7			
AAB90886				
ID	AAB90886	standard;	Peptide; 7	AA.
XX				
XX	AAB90886;			
XX				
DT	22-JUN-2001	(first entry)		
XX				
DE	Vasopressin (ADH)	related peptide	SEQ ID NO:58.	
XX				
KX	Protection; endogenous therapeutic peptide;	peptidase; conjugation;		
KW	blood component; modification; succinimidyl;	maleimido group; amino;		
KW	hydroxyl; thiol; hormone; growth factor;	neurotransmitter.		
XX				
OS	Homo sapiens.			
OS	Synthetic.			
XX				
PN	MO200069900-A2.			
XX				
PD	23-NOV-2000.			
XX				
XX				
PE	17-MAY-2000.	2000WO-US13576.		

XX	17-MAY-1999;	9905-0134406.
XX	PR 10-SEP-1999;	9905-0153406.
XX	PR 15-OCT-1999;	9905-0159783.
XX		
XX	(CONJ-) CONJUCHEM INC.	
XX		
XX	Bridon DP, Ezrin AM, Milner PG, Holmes DL, Thibaudau K:	
XX	WPI; 2001-112059/12.	
XX		
XX	Modifying and attaching therapeutic peptides to albumin prevents	
XX	peptidase degradation, useful for increasing length of in vivo activity	
XX	-	
XX		
XX	Disclosure; Page 208; 73pp; English.	
XX		
XX	The present invention describes a modified therapeutic peptide (I)	
XX	comprising a therapeutically active amino acid region (III) and a	
XX	reactive group (II) (e.g. succinimidy and maleimido groups) attached to	
XX	a less therapeutically active amino acid region (IV), which covalently	
XX	bonds with amino/hydroxyl/thiol groups on blood components to form a	
XX	peptidase stabilised therapeutic peptide composed of 3-50 amino acids.	
XX	(I) are useful for modifying therapeutic peptides e.g. hormones, growth	
XX	factors and neurotransmitters, to protect them from peptidase activity	
XX	in vivo for the treatment of various disorders. Endogenous therapeutic	
XX	peptides are not suitable as drug candidates as they require frequent	
XX	administration due to rapid degradation by peptidases in the body.	
XX	Modifying and attaching therapeutic peptides to albumin prevents or	
XX	reduces the action of peptidases to increase length of activity (half	
XX	life) and specificity as bonding to large molecules decreases	
XX	intracellular uptake and interference with physiological processes.	
XX	AAB90829 to AAB92441 represent peptides which can be used in the	
XX	exemplification of the present invention.	
XX		
XX	Sequence 7 AA:	
XX		
XX	Query Match 9.6%; Score 5; DB 22; Length 7;	
XX	Best local Similarity 100.0%; Pred. No. 4.3e+05;	
XX	Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

QY	3	VNCPG 7	
Db	3	vncpg 7	
	RESULT 8		
	AAB90896		
ID	AAB90896	standard; Peptide; 7 AA.	
XX			
XX	AAB90896;		
AC			
XX			
DT	22-JUN-2001	(first entry)	
XX			
DE	Vasopressin (ADH) related peptide SEQ ID NO:68.		
XX			
KW	Protection; endogenous therapeutic peptide; peptidase; conjugation;		
KW	blood component; modification; succinimidyl; maleimido group; amino;		
XX	hydroxyl; thiol; hormone; growth factor; neurotransmitter.		
XX			
OS	Homo sapiens.		
OS	Synthetic.		
XX			
PN	WO200069900-A2.		
XX			
XX			
PD	23-NOV-2000.		
XX			
PF	17-MAY-2000; 2000WO-US13576.		
XX			
PR	17-MAY-1999; 99US-0134406.		
PR	10-SEP-1999; 99US-0153406.		
PR	15-OCT-1999; 99US-0159783.		





Best Local Similarity 100.0%; Pred. No. 43e+05;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 PENGs 45  
|||||  
Db 5 pengs 9

RESULT 11  
AAV06014  
ID AAV06014 standard; Peptide; 10 AA.

XX  
AC AAV06014;

XX 16-AUG-1999 (first entry)

XX Human cancer antigen NY ESO-1/CAG-3 HLA peptide motif.

XX NY ESO-1/CAG-3 gene; cancer peptide; antigen; human;  
XX leukemia; non-Hodgkins lymphoma; Hodgkins lymphoma; lung cancer;  
XX metastasis; melanoma; adenocarcinoma; thymoma; colon cancer;  
XX uterine cancer; breast cancer; prostate cancer; ovarian cancer;  
XX cervical cancer; bladder cancer; kidney cancer; pancreatic cancer;  
XX liver cancer; sarcoma; tumour; diagnosis; immunotherapy; therapy;  
XX vaccine; human leukocyte antigen; HLA.

XX Homo sapiens.

XX W09918206-A2.

XX 15-APR-1999.

XX 21-SEP-1998; 98WO-US19609.

XX 08-OCT-1997; 97US-0061428.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Rosenberg SA, Wang RF;

XX WPI; 1999-277270/23.

XX Cancer antigen NY ESO1/CAG-3

XX Example 10; Page 42; 88pp; English.

XX This peptide was identified as an HLA peptide motif following a  
XX screen for epitopes from the coding region of human ESO-1/CAG-3  
XX ORF1 (see AX58599). 30 Epitopes (see AAY05988-T06017) were identified.  
XX The present peptide (ranked 27) corresponds to amino acid residues  
XX 22-31 of CAG-1 ORF1 (see AAY05965). CAG-1 is a new and potent  
XX tumour antigen capable of eliciting an antigen specific immune  
XX response by T cells. Cancer peptides (see AAY05967-87) derived from  
XX CAG-3, portions of CAG-3 and their variants, are useful as cancer  
XX vaccines. A claimed method of preventing or inhibiting cancer  
XX involves administering a cancer peptide, with or without an HLA  
XX molecule. The cancer peptides form part of, or are derived  
XX from, cancers such as primary or metastatic melanoma, thymoma,  
XX lymphoma, sarcoma, lung cancer, liver cancer, leukaemia, uterine  
XX cancer, cervical cancer, bladder cancer, kidney cancer and  
XX adenocarcinomas such as breast, prostate, ovarian, pancreatic and  
XX thyroid cancers.

XX Sequence 10 AA;

Query Match 9.6%; Score 5; DB 20; Length 10;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PDGPG 52  
|||||  
Db 3 pdpgp 7

RESULT 12

AAAG97571  
ID AAG97571 standard; Peptide; 10 AA.

XX  
AC AAG97571;

XX 18-SEP-2001 (first entry)

XX Human complementary peptide, SEQ ID NO: 3766.

XX Human; complementary peptide; ligand; drug discovery; drug design.

XX Homo sapiens.

XX W0200142277-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04776.

XX 13-DEC-1999; 99GB-0029464.

XX (PROT-) PROTEOM LTD.

XX Roberts GW, Heal JR;

XX WPI; 2001-408419/43.

XX A set of peptide ligands consisting of specific complementary peptides  
XX to proteins encoded by genes of the human genome, useful in an assay  
XX for screening and identifying of one or more novel peptides which are  
XX drug candidates or pro-drugs -

XX Example 6; Page 583; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
XX generated from the human genome. The complementary peptides  
XX interact with their relevant target proteins encoded in the human  
XX genome. They can be used as reagents in drug discovery and as lead  
XX ligands to facilitate drug design and development. The present  
XX sequence is a complementary peptide provided in the specification.

XX Sequence 10 AA;

Query Match 9.6%; Score 5; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 GGINA 11  
|||||  
Db 3 ggina 7

RESULT 13

AAAG97572  
ID AAG97572 standard; Peptide; 10 AA.

XX  
AC AAG97572;

XX 18-SEP-2001 (first entry)

XX Human complementary peptide, SEQ ID NO: 3767.

XX Human; complementary peptide; ligand; drug discovery; drug design.

XX Homo sapiens.

XX W0200142277-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04776.  
PF 13-DEC-1999; 99GB-0029464.  
PR (PROT-) PROTEOM LTD.  
PA Roberts GW, Heal JR;  
PI WPI; 2001-367863/38.  
DR WPI; 2001-408419/43.  
XX  
PT A set of peptide ligands consisting of specific complementary peptides  
PT to proteins encoded by genes of the human genome, useful in an assay  
PT for screening and identifying of one or more novel peptides which are  
PT drug candidates or pro-drugs -  
XX  
PS Example 6; Page 583; 646pp; English.  
XX  
CC The invention relates to a set of complementary peptide ligands  
CC generated from the human genome. The complementary peptides  
CC interact with their relevant target proteins encoded in the human  
CC genome. They can be used as reagents in drug discovery and as lead  
CC ligands to facilitate drug design and development. The present  
CC sequence is a complementary peptide provided in the specification.  
XX  
SQ Sequence 10 AA;

Query Match 9.6%; Score 5; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 7 GGINA 11  
      |||||  
Db 3 ggina 7

RESULT 14  
AAG87784  
ID AAG87784 standard; Peptide; 10 AA.  
XX  
AC AAG87784;  
XX  
DT 11-SEP-2001 (first entry)  
XX  
DE Saccharomyces cerevisiae peptide; SEQ ID NO: 2733.  
XX  
KW Saccharomyces cerevisiae; complementary peptide; peptide identification;  
KW drug discovery; drug design.  
XX  
OS Saccharomyces cerevisiae.  
XX  
PN WO200142276-A1.  
XX  
PD 14-JUN-2001.  
XX  
PF 13-DEC-2000; 2000WO-GB04773.  
XX  
PR 13-DEC-1999; 99GB-0029471.  
XX  
PA (PROT-) PROTEOM LTD.  
XX  
PI Roberts GW, Heal JR;  
XX  
DR WPI; 2001-367863/38.  
XX  
PT Identifying complementary peptides by analysis of protein and  
PT nucleotide sequence databases, useful in drug design -  
XX  
PS Example 5; Page 407; 488pp; English.  
XX  
CC The invention relates to the identification of complementary peptides  
CC by analysis of protein and nucleotide sequence databases from higher  
CC eukaryotic genomes, excluding human and plants. The specific  
CC complementary peptides interact with their relevant target proteins  
CC encoded in the eukaryotic genome. The peptides may be used as reagents  
CC and drugs in drug discovery and as lead ligands for drug design and  
CC development. The present sequence is a complementary peptide from  
CC Saccharomyces cerevisiae.  
XX  
SQ Sequence 10 AA;

CC eukaryotic genomes, excluding human and plants. The specific  
CC complementary peptides interact with their relevant target proteins  
CC encoded in the eukaryotic genome. The peptides may be used as reagents  
CC and drugs for drug discovery and as lead ligands for drug design and  
CC development. The present sequence is a complementary peptide from  
CC Saccharomyces cerevisiae.  
XX  
SQ Sequence 10 AA;

Query Match 9.6%; Score 5; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 7 GGINA 11  
      |||||  
Db 3 ggina 7

RESULT 15  
AAG87785  
ID AAG87785 standard; Peptide; 10 AA.  
XX  
AC AAG87785;  
XX  
DT 11-SEP-2001 (first entry)  
XX  
DE Saccharomyces cerevisiae peptide; SEQ ID NO: 2734.  
XX  
KW Saccharomyces cerevisiae; complementary peptide; peptide identification;  
KW drug discovery; drug design.  
XX  
OS Saccharomyces cerevisiae.  
XX  
PN WO200142276-A1.  
XX  
PD 14-JUN-2001.  
XX  
PF 13-DEC-2000; 2000WO-GB04773.  
XX  
PR 13-DEC-1999; 99GB-0029471.  
XX  
PA (PROT-) PROTEOM LTD.  
XX  
PI Roberts GW, Heal JR;  
XX  
DR WPI; 2001-367863/38.  
XX  
PT Identifying complementary peptides by analysis of protein and  
PT nucleotide sequence databases, useful in drug design -  
XX  
PS Example 5; Page 407; 488pp; English.  
XX  
PD The invention relates to the identification of complementary peptides  
PD by analysis of protein and nucleotide sequence databases from higher  
PD eukaryotic genomes, excluding human and plants. The specific  
PD complementary peptides interact with their relevant target proteins  
PD encoded in the eukaryotic genome. The peptides may be used as reagents  
PD and drugs in drug discovery and as lead ligands for drug design and  
PD development. The present sequence is a complementary peptide from  
PD Saccharomyces cerevisiae.  
XX  
SQ Sequence 10 AA;

Query Match 9.6%; Score 5; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 7 GGINA 11  
      |||||  
Db 3 ggina 7

Thu Mar 28 09:21:34 2002

us-09-726-348-2\_copy\_126\_177.rag

Search completed: March 28, 2002, 09:15:55  
Job time: 854 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 07:32:33 ; Search time 105.47 Seconds  
(without alignments)  
160.830 Million cell updates/sec

Title: US-09-726-348-2

Perfect score: 1243

Sequence: 1 MAPHGPGSLTLVPMVAALL.....TTLVSILLMATORKKAKTS 229

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SIDSI/gcgdata/geneseq/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1243	100.0	229	18	AAW27087
2	1243	100.0	229	20	AAV13944
3	1243	100.0	229	21	AAH33419
4	1243	100.0	229	21	AAV88570
5	1243	100.0	229	22	AAU04295
6	1243	100.0	229	22	AAH20112
7	1243	100.0	229	22	AAH68595
8	1243	100.0	229	22	AAH50951
9	1243	100.0	229	22	AAH48068
10	1231	99.0	229	20	AAH05282
11	790	63.6	147	21	AAH00157

12	303	24.4	52	22	AAU04296
13	103	8.3	173	22	AAW24243
14	103	8.3	383	22	AAE01167
15	102.5	8.2	1404	21	AAV59600
16	101	8.1	1964	20	AAW55557
17	99	8.0	298	21	AAH07469
18	99	8.0	769	16	AAH75352
19	98.5	7.9	1218	18	AAH18354
20	98	7.9	670	16	AAH67759
21	97.5	7.8	1712	13	AAH22461
22	97	7.8	1008	22	AAH82247
23	96.5	7.8	2199	17	AAH45562
24	96.5	7.8	2201	22	AAH36935
25	96.5	7.8	2471	20	AAH06816
26	96	7.7	880	21	AAH01249
27	96	7.7	886	22	AAH71869
28	96	7.7	1193	17	AAH05835
29	96	7.7	1193	21	AAH59599
30	94.5	7.6	281	22	AAU12198
31	93.5	7.5	2157	21	AAH93910
32	93.5	7.5	3298	22	AAH03657
33	93	7.5	1050	22	AAH66267
34	92.5	7.4	385	15	AAH60176
35	92.5	7.4	1010	20	AAH87896
36	92.5	7.4	1036	18	AAH18351
37	92.5	7.4	1187	18	AAH18352
38	92.5	7.4	1208	19	AAH40827
39	92.5	7.4	1218	17	AAH05833
40	92.5	7.4	1218	19	AAH44301
41	92.5	7.4	1218	20	AAH87894
42	92.5	7.4	1218	21	AAH59597
43	91.5	7.4	1404	14	AAH38304
44	91.5	7.4	751	15	AAH53088
45	91.5	7.4	752	15	AAH53087

#### ALIGNMENTS

RESULT 1	
ID	AAW27087
AAW27087	standard; Protein; 229 AA.
AC	AAW27087;
DT	28-JAN-1998 (first entry)
DE	Human transforming growth factor alpha HIII.
XX	
XX	human transforming growth factor; TGF; TGF-alpha-HIII; angiogenesis;
KW	embryogenesis; ocular disorder; kidney disorder; liver disorder;
KW	neural disorder; alopecia; inflammation.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..25
FT	/label= signal_peptide
FT	/note= "putative"
FT	26..229
FT	/label= mature_protein
FT	1..177
FT	/label= soluble_portion_of_protein
FT	126..177
FT	/label= active_site
FT	178..204
FT	/label= transmembrane_portion
FT	/note= "putative"
XX	
XX	W09725349-A1.
XX	17-JUL-1997.
XX	

Transforming growth factor beta encoded protein  
Human gene 4 encodes transforming growth factor beta  
Drosophila Serrate  
Mus musculus notch  
A human leucine-rich repeat domain  
Human fetal brain  
Proliferation and differentiation  
Human fetal brain  
Masking protein h1  
Rat insulin-respons  
Human cytotactin.  
Human tenascin-C.  
Human Notch2 (human)  
Human EMR1 hormone  
Human EMR1 seven t  
Chick Serrate. Ga  
Chick Serrate prot  
Human PRO1341 poly  
A human hyaluronan  
Human extracellular  
Human TANGO 272 SE  
Chimeric protein (Human JAGGED1 solu  
Proliferation and  
Human Jagged prote  
Human Serrate-1 (H  
Human Serrate 1.  
Human JAGGED1 prot  
Human Serrate prot  
Sequence of a ser  
Human masking prot  
Human masking prot

PF 04-JAN-1996; 96WO-US00149.  
 XX  
 XX 04-JAN-1996; 96WO-US00149.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX  
 PI Wei Y;  
 DR WPI: 1997-372817/34.  
 N-PSDB; AAT85082.  
 XX  
 XX New human transforming growth factor-alpha homologue - used for  
 PT developing products for treating e.g. neurological disorders, kidney  
 PT and liver disorders, tumours, wounds, hair loss or skin disorders  
 PS  
 PS Claim 15; Page 47; 63pp: English.  
 XX  
 XX This protein has been putatively identified as a human transforming  
 CC growth factor (TGF) alpha analogue, TGF-alpha-H11. The protein can  
 CC stimulate angiogenesis, embryogenesis, cell differentiation and function.  
 CC It can be used for therapeutic purposes for restoration or enhancement of  
 CC neurological functions diminished as a result of trauma or other damaging  
 CC pathologies such as AIDS dementia and senile dementia, to treat ocular  
 CC disorders, e.g. corneal inflammation, to destroy target cells, to treat  
 CC tumours, kidney or liver disorders or to treat wounds, burns or ulcers.  
 CC The polypeptide can also be used in the modulation of angiogenesis, bone  
 CC resorption, immune response, and synaptic and neuronal effector  
 CC functions, or the arachidonic acid cascade. It can also be used in  
 CC applications related to terminal differentiation e.g. in  
 CC hyperproliferative disorders such as inflammation or psoriasis and for  
 CC alopecia, hair loss or other skin conditions which affect hair follicular  
 CC development. Antagonists to TGF-alpha-H11 can be used for treating  
 CC tumours or skin disorders such as psoriasis. The products can also  
 CC be used for diagnosis and detection of the above disorders.  
 CC  
 SO Sequence 229 AA;

Query Match 100.0%; Score 1243; DB 18; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 2,1e-102;  
 Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAHGPESLITLVPMMAALLALGVERALALPEICTQCPGSVQNSKVAFTCKTTRRLM 60  
 DB 1 maphgpslittlvpmmaallalgyeralalpeictqcpgsvqnskvaftcktrrelm 60  
 OY 61 HARCCLMQKGTITGLDLQNSLEDPGNFHQAHTVYIDLANPKGLDANTFRGTQLQ 120  
 DB 61 harccclmqgtitglldlqnsledpgnfhqahvtvtyidlanpklgdiantfrgtqlq 120  
 OY 121 TLIPQHVNCPGGINAMNTTSTYIDNOICOGKNCNNTGDEMPCEPENGSCVPDGPGLQ 180  
 DB 121 tllipqhvnpcpgginamnttstydnoicogkncnntgdepmpcengscvpdgpqllq 180  
 OY 181 CVCADGFHGYKCMRQGSFSLMFFGLGATTLVSILLMAOTRRAKTS 229  
 DB 181 cvcadgfhygkcmrqsfsllmffglgattllvsillwatqrrakts 229

RESULT 2  
 AAY13944  
 ID AAY13944 standard; Protein: 229 AA.  
 AC AAY13944;  
 XX  
 XX 14-JUL-1999 (first entry)  
 DT  
 XX  
 DE Human transmembrane protein, HP10435.  
 XX  
 XX Transmembrane protein: human; cell membrane; proliferation; diagnosis;  
 KM cell differentiation; carcinostatic agent; probe; gene therapy;  
 KW signal transduction; apoptosis; inhibitor;  
 KW phosphatidylincholine N-methyltransferase.

XX Homo sapiens.  
 OS  
 XX  
 XX WC9918203-A2.  
 PN  
 XX  
 XX 15-APR-1999.  
 PD  
 XX  
 XX 05-OCT-1998; 98WO-JP04475.  
 PF  
 XX  
 XX 08-OCT-1997; 97JP-0276271.  
 PR  
 XX  
 XX (PROT-) PROTEGENE INC.  
 PA (SAGA) SAGAMI CHEM RES CENT.  
 XX  
 XX Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
 PI  
 DR WPI: 1999-277268/23.  
 N-PSDB; AAX36812, AAX36813.  
 XX  
 XX Human transmembrane proteins and nucleotide sequences  
 PS  
 PS Claim 1; Page 90-91; 139pp: English.  
 XX  
 XX This sequence is a human transmembrane protein of the invention.  
 CC All of the proteins exist in the cell membrane, so are considered to be  
 CC proteins controlling the proliferation and differentiation of the cells.  
 CC They may be useful as carcinostatic agents or as antigens for preparing  
 CC antibodies against the proteins. The CDNA can be used as probes for  
 CC gene diagnosis and gene sources for gene therapy, as well as for  
 CC large-scale expression of the proteins. The HP01498 (see AAY13939)  
 CC protein may be associated with signal transduction associated with  
 CC apoptosis, and therefore useful in inhibition of apoptosis. The HP01962  
 CC (see AAY13943) protein can be used to treat diseases associated with  
 CC phosphatidylincholine N-methyltransferase. The proteins are  
 CC identified by the presence of a hydrophobic transmembrane region,  
 CC knowledge of the protein function is not required, as in e.g. methods of  
 CC expression cloning.  
 CC  
 SO Sequence 229 AA;

Query Match 100.0%; Score 1243; DB 20; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 2,1e-102;  
 Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAHGPESLITLVPMMAALLALGVERALALPEICTQCPGSVQNSKVAFTCKTTRRLM 60  
 DB 1 maphgpslittlvpmmaallalgyeralalpeictqcpgsvqnskvaftcktrrelm 60  
 OY 61 HARCCLMQKGTITGLDLQNSLEDPGNFHQAHTVYIDLANPKGLDANTFRGTQLQ 120  
 DB 61 harccclmqgtitglldlqnsledpgnfhqahvtvtyidlanpklgdiantfrgtqlq 120  
 OY 121 TLIPQHVNCPGGINAMNTTSTYIDNOICOGKNCNNTGDEMPCEPENGSCVPDGPGLQ 180  
 DB 121 tllipqhvnpcpgginamnttstydnoicogkncnntgdepmpcengscvpdgpqllq 180  
 OY 181 CVCADGFHGYKCMRQGSFSLMFFGLGATTLVSILLMAOTRRAKTS 229  
 DB 181 cvcadgfhygkcmrqsfsllmffglgattllvsillwatqrrakts 229

RESULT 3  
 AAB33419  
 ID AAB33419 standard; Protein: 229 AA.  
 AC AAB33419;  
 XX  
 XX 29-JAN-2001 (first entry)  
 DT  
 XX  
 DE Human PR0240 protein UNQ214 SEQ ID NO:26.  
 XX  
 XX Human; immune related disease; diagnosis; antiinflammatory; cardiant;

dermatological; antiarthritic; antirheumatic; immunosuppressive;  
 haemostatic; antithyroid; antidiabetic; noctropic; neuroprotective;  
 antianemic; hepatotropic; vitruide; antiporiatic; antiallergic;  
 antiastrumatic; systemic lupus erythematosus; rheumatoid arthritis;  
 osteoarthritis; spondyloarthropathy; systemic sclerosis; sarcoidosis;  
 idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;  
 systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;  
 autoimmune thrombocytopaenia; immune-mediated renal disease;  
 demyelinating disease; hepatobiliary disease; Whipple's disease;  
 inflammatory bowel disease; gluten sensitive enteropathy;  
 autoimmune disease; immune-mediated skin disease; allergic disease;  
 immunological disease; transplantation associated disease;  
 graft rejection; graft-versus-host-disease.

Homo sapiens.

WO200053758-A2.

14-SEP-2000.

02-MAR-2000; 2000WO-US05841.

08-MAR-1999; 99WO-US05028.  
 10-MAR-1999; 99US-0123618.  
 12-MAR-1999; 99US-0123957.  
 23-MAR-1999; 99US-0125775.  
 12-APR-1999; 99US-0128849.  
 20-APR-1999; 99WO-US08615.  
 28-APR-1999; 99US-0131445.  
 04-MAY-1999; 99US-0132371.  
 14-MAY-1999; 99US-0134287.  
 02-JUN-1999; 99WO-US12252.  
 23-JUN-1999; 99US-0141037.  
 20-JUL-1999; 99US-0144758.  
 26-JUL-1999; 99US-0145698.  
 28-JUL-1999; 99US-0146222.  
 01-SEP-1999; 99WO-US20111.  
 08-SEP-1999; 99WO-US20594.  
 13-SEP-1999; 99WO-US20944.  
 15-SEP-1999; 99WO-US21090.  
 15-SEP-1999; 99WO-US21547.  
 05-OCT-1999; 99WO-US23089.  
 29-OCT-1999; 99US-0162506.  
 29-NOV-1999; 99WO-US28214.  
 30-NOV-1999; 99WO-US28313.  
 30-NOV-1999; 99WO-US28409.  
 01-DEC-1999; 99WO-US28301.  
 01-DEC-1999; 99WO-US28634.  
 02-DEC-1999; 99WO-US28551.  
 02-DEC-1999; 99WO-US28564.  
 16-DEC-1999; 99WO-US28565.  
 20-DEC-1999; 99WO-US30095.  
 16-DEC-1999; 99WO-US30999.  
 30-DEC-1999; 99WO-US31274.  
 05-JAN-2000; 2000WO-US00217.  
 06-JAN-2000; 2000WO-US00376.  
 11-FEB-2000; 2000WO-US03565.  
 18-FEB-2000; 2000WO-US04342.  
 22-FEB-2000; 2000WO-US04414.

(GETH ) GENENTECH INC.

Ashtkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W;  
 Kabakoff RC, Lu Y, Pan J, Pennica D, Shelton DL, Smith V;  
 Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M;  
 WPI; 2000-572271/53.  
 N-PSDB; AAC58584.

Sixty four PRO polypeptides, useful in the diagnosis and treatment of  
 immune related disorders, e.g. systemic lupus erythematosus, rheumatoid

arthritis, osteoarthritis, thyroiditis and diabetes mellitus -

Claim 33; Fig 12; 309pp; English.

The present invention describes sixty four human PRO proteins which can  
 be used in the treatment of immune related diseases. The human PRO  
 proteins, anti-PRO antibodies, agonists and antagonists are useful for  
 treating and diagnosing immune related disorders. The disorders are  
 selected from systemic lupus erythematosus, rheumatoid arthritis,  
 osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,  
 systemic sclerosis, idiopathic chronic arthritis, Sjogren's  
 syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic  
 anaemia, autoimmune thrombocytopaenia, thyroiditis, diabetes mellitus,  
 immune-mediated renal disease, demyelinating diseases of the central  
 and peripheral nervous systems, hepatobiliary diseases, inflammatory  
 bowel disease, gluten-sensitive enteropathy and Whipple's disease,  
 autoimmune or immune-mediated skin diseases, allergic diseases,  
 immunological diseases of the lung, and transplantation associated  
 diseases including graft rejection and graft-versus-host-disease.  
 AAC58397 to AAC58578 represent PCR primers and hybridisation probes used  
 in the isolation of human PRO sequences. AAC58579 to AAC58642 and  
 AAC583414 to AAC583477 represent human PRO polynucleotide and protein  
 sequences given in the exemplification of the present invention.

Sequence 229 AA:

Query Match 100.0%; Score 1243; DB 21; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 2,1e-102;  
 Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPHGSGSLTTLVPAAALILALIGERALALPEICTQCGSVONLSKYAFCKTRELML 60  
 DB 1 maphgsgsltlvpwaaalilalgyeralalpeictcgpsvgnskvafcktreml 60  
 QY 61 HARCCLNOKGTILGLDIONCSLEDPGNFHOAHNTVIIDQANPLKGLUATFRFQIQ 120  
 DB 61 harcclngkgtlglidngcsledpgpnfhqahnttcvllidqanplkgldatfrfqlq 120  
 QY 121 TLLPQHVNCPSGGINAMWNTTSYIDMOTGCKNLGNNTGPEKPEKSCVPDPGGLIQ 180  
 DB 121 tllpqhvnpcpgginawnttsyidngicgqgknlcntgdepemcngscvpdpqglilq 180  
 QY 181 CVCADGFHGYCKMGSGSFLMFFGLGATTVSVILIMATQRRKAKTS 229  
 DB 181 cvcadgfhgycmrgsgsflmffglgatlsvsllwatqrrkakts 229

RESULT 4  
 AAY88570  
 ID AAY88570 standard; Protein; 229 AA.  
 AC AAY88570;  
 XX  
 DT 09-AUG-2000 (first entry)  
 XX  
 DE Human PRO240 amino acid sequence.  
 XX  
 KW Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;  
 KW PRO117; tumour growth inhibitor; cancer; diagnosis; treatment; human;  
 KW cell growth proliferation; serrate precursor; C-serrate-1; ADPPT;  
 KW antibody dependent enzyme mediated produg therapy; chromosome 2.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200015666-A2.  
 XX  
 PD 23-MAR-2000.  
 XX  
 PF 08-SEP-1999; 99WO-US20594.  
 XX  
 PF 10-SEP-1998; 98US-0099803.  
 XX  
 PR 10-SEP-1998; 98WO-US18824.

XX (GENE ) GENE/TECH INC.  
 PA  
 XX  
 PI Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WT, Botstein D;  
 XX  
 XX WPI: 2000-271386/23.  
 DR N-PSDB: AAA30036.

XX New isolated antibodies which bind to specific polypeptides used for  
 PT diagnosis and treatment of neoplastic cell growth and proliferation -  
 XX

PS Example 4; Fig 8; 200pp; English.

XX This sequence represents a human PRO240 amino acid sequence. PRO240  
 CC shares sequence homology with the D. melanogaster serrate precursor  
 CC protein and the gallus gallus C-serrate-1 protein. The PRO240 gene is  
 CC located on chromosome 2. The invention relates to isolated antibodies  
 CC which bind to a polypeptide. The "PRO" polypeptides are encoded by genes  
 CC which are over expressed in the genome of tumour cells. Vectors and host  
 CC cells comprising the nucleic acid encoding the antibodies are used in the  
 CC production of the antibodies. The antibodies and nucleic acids encoding  
 CC them are used for diagnosing a tumour in a mammal. The antibodies are  
 CC used for inhibiting the growth of tumour cells and identifying compounds  
 CC that inhibit a biological or immunological activity of and/or expression  
 CC of a PRO187, PRO533, PRO214, PRO240, PRO211, PRO230, PRO261, PRO246 or  
 CC PRO317 polypeptide. The antibody can be used in antibody dependent enzyme  
 CC mediated prodrug therapy (ADEPT) by conjugating the antibody to a  
 CC prodrug-activating enzyme which converts a prodrug to an anti-cancer  
 CC drug. The antibodies can be fluorescently labelled and monitored by light  
 CC microscopy, flow cytometry or fluorimetry for diagnosis and prognosis of  
 CC tumours.

XX Sequence 229 AA:

Query Match 100.0%; Score 1243; DB 21; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-102;  
 Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPHGPGSLTTLVMAAALLALGVERALALPEICTQCQSGSVONLSKVAFYCKTTRRLML 60  
 Db 1 maphgpgsltlvmaaaallalgyeralalpeictcpgsvgnlskvaifyckttreml 60  
 OY 61 HARCLNKGKTIIGLDQNGSLDPGPFHQATHTYIIDQANPLKGDLANTFRGFTQLQ 120  
 Db 61 harclnkgktilgldqngslpdpghqathtvtiidaqnpkldantfrgtftq 120  
 OY 121 TLIPQHVNCFGINAMNTTSYIDNQCQGNLCNNTGDPMPCEPENGSCVPDGPGLQ 180  
 Db 121 tllpqhvnncpginamnttstydncqgknlcnntgdpmpcengscvpdgp 180  
 OY 181 CVCADGFRHKGKCMQSGFSLLMPFGILGATTSVITLMAOTRRKAKTS 229  
 Db 181 cvcadgfrhkgkcmqsgfsllmftfgilgattlsvitlwatqrrakts 229

RESULT 5

AAU04295 standard; Protein: 229 AA.

AAU04295:

24-OCT-2001 (first entry)

Transforming growth factor (TGF) alpha HIII.

Human; TGF alpha HIII; transforming growth factor alpha HIII; cancer;  
 diagnostic; therapeutic; immune disorder; multiple sclerosis;  
 systemic lupus erythematosus; human immuno-deficiency virus; HIV;  
 hyperproliferative disorder; Gaucher's disease; cardiovascular disease;  
 Schmltar syndrome; Chaga's cardiomyopathy; coronary arteriosclerosis;  
 angiogenic disorder; corneal graft; neovascularisation; wound healing;  
 diabetic retinopathy; neurological disorder; Huntington's chorea;

Alzheimer's disease; Parkinson's disease.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..25

Protein /label= signal\_peptide

Active-site /note= "Mature TGF alpha HIII"

Region 126..177

178..204

/note= "transmembrane region"

01-DEC-2000; 2000WO-US32745.

02-DEC-1999; 99US-0168387.

(HUMA-) HUMAN GENOME SCI INC.

WPI: 2001-441480/47.

N-PSDB: AAS08543.

Nucleic acid encoding human transforming growth factor alpha HIII

(TGF $\alpha$ ), useful for preventing, diagnosing and/or treating e.g. Cancer

and Parkinson's disease -

Claim 11; Fig 1; 302pp; English.

The sequence represents the amino acid sequence of human transforming  
 growth factor (TGF) alpha HIII. TGF alpha HIII nucleic acid and protein  
 may be used in the prevention, diagnosis and treatment of diseases  
 associated with inappropriate polypeptide expression, for example immune  
 disorders (e.g. multiple sclerosis, systemic lupus erythematosus and  
 human immuno-deficiency virus (HIV) infections), hyperproliferative  
 disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases  
 (e.g. Schmltar syndrome, Chaga's cardiomyopathy and coronary  
 arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 neovascularisation) and diabetic retinopathy, neurological disorders  
 (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 infectious diseases and/or for promoting wound healing, regeneration  
 and/or chemotaxis (full details given in specification). Additionally,  
 the nucleic acid may be used to produce the secreted polypeptides, by  
 inserting the nucleic acids into a host cell and culturing the cell to  
 express the protein. It may also be used as a DNA probe in diagnostic  
 assays to detect and quantitate the presence of similar nucleic acid  
 sequences in samples, and therefore which patients may be in need of  
 restorative therapy. The polypeptides may also be used as antigens in the  
 production of antibodies against TGF alpha HIII and in assays to identify  
 modulators of TGF alpha HIII. The anti-TGF alpha HIII antibodies may also  
 be used as diagnostic agents for detecting the presence of TGF alpha HIII  
 in samples (e.g. by enzyme linked immunosorbant assay (ELISA)).

Sequence 229 AA:

Query Match 100.0%; Score 1243; DB 22; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-102;  
 Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MAPHGPGSLTTLVMAAALLALGVERALALPEICTQCQSGSVONLSKVAFYCKTTRRLML 60  
 Db 1 maphgpgsltlvmaaaallalgyeralalpeictcpgsvgnlskvaifyckttreml 60  
 OY 61 HARCLNKGKTIIGLDQNGSLDPGPFHQATHTYIIDQANPLKGDLANTFRGFTQLQ 120  
 Db 61 harclnkgktilgldqngslpdpghqathtvtiidaqnpkldantfrgtftq 120



PI Hillan KJ, Mark MR, Maisters SA, Pitti RM, Tumas D, Watanabe CK;  
wood wi;

Db 181 cvcadgfnhykcmrgsfsllmffgilgatlsvsillwatqrrkaks 229

blastocoe

The present sequence is one of twenty eight novel PRO polypeptides. The PRO polypeptides and their agonists, including antibodies, peptides, and small molecule agonists, may be used to treat various tumours, e.g., cancers such as breast cancer, ovarian cancer, renal cancer, colorectal cancer, uterine cancer, prostate cancer, lung cancer, bladder cancer, central nervous system cancer, melanoma or leukaemia. They are also useful for treating other disorders such as neuronal, glial, astrocytal and hypothalamic and other glandular, macrophagal, epithelial, stromal and fibrocellular disorders, and inflammatory, angiogenic and immunological disorders.

XX	SEQUENCE	229 AA:
CC	Query Match	100.0%; Score 1243; DB 22; Length 229;
CC	Best Local Similarity	100.0%; Pred. No. 2.le-102;
CC	Matches 229; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
YY	1 MPHGGSLTTLTPWMAALLLALGVERALALEICTCQCGSQVNSKVAFYCKTRLEML	60
DB	1 mephpgsstlttlvpwaaalllalyeralalpeltctqcpssvqlskvalfycktreml	60
OY	61 HARCCLNOKGTLLGLDQLQCNSLEDPCGNFQAHTTVIIDLANPKGDIAHFRTPTOL	120
DB	61 harcclnqggtlllgldlqcnsledpcgnfnghattvtidlanpkgdianfrfttqlq	120
OY	122 TLILQHVNCPGGINAANNITSTYINDQICQGKNLCNNTGDEPCKPENGSCVPDGPGLIQ	180
DB	121 tlllpqhvnccpgginannitstytindiqcgqknlcnntgdepengscvdpdgpjllq	180
OY	181 CVCADGFHGFKCMRGSGFSLLMFEGILLGATTSVILLMATORRAKTS	229
DB	181 cvcadgfingfkcmrgsfelllmfifglgatlsvsillmatgrakts	229
RESULT	9	
ID	AAB48068	
AC	AAB48068 standard; protein; 342 AA.	
AA	AAB48068;	
DT	19-MAR-2001 (first entry)	
DE	Human extracellular signaling molecule (EXCS) (ID 2207183CD1).	
KW	Extracellular signaling molecule; EXCS; anti-inflammatory; human;	
KW	immunosuppressive; cytosolic; neuroprotective; gastrointestinal;	
KW	viral; antibacterial; anti-HIV; human immunodeficiency virus;	
KW	antiferility; cerebroprotective; nootropic; antifungal; antifungal;	
KW	anticonvulsant; tranquillizer; neuroleptic; vasotrophic; gynecological;	
KW	keratolytic; protozoacide; gene therapy.	
OS	Homo sapiens.	
PN	WO200070049-A2.	
PD	23-NOV-2000.	
PF	19-MAY-2000; 2000WO-US13975.	
PR	19-MAY-1999; 99US-0134949.	
PR	15-JUL-1999; 99US-0144270.	
PR	30-JUL-1999; 99US-0146700.	
PR	04-OCT-1999; 99US-0157508.	
PA	(INCY-) INCYTE GENOMICS INC.	
PI	Tang YT, Yue H, Lal P, Burford N, Bandman O, Baughn MR,	
PI	Azizmai Y, Lu DAM, Patterson C;	
DR	WPI: 2001-025021/03.	
N-PSDB:	AAC84304.	
PT	New human extracellular signaling nucleic acids and polypeptides useful	
PT	for diagnosing, treating and preventing infections and	
PT	gastrointestinal, neurological, reproductive, and	
PT	autoimmune/inflammatory disorders -	
Claim 1:	Page 89-90; 114pp; English.	
The invention provides human extracellular signaling molecules (EXCS)		
and polynucleotides which identify and encode EXCS. EXCS can be		
expressed by standard recombinant methodology. The amino acid and nucleic		
acid sequences of EXCS are useful for diagnosing, treating and		

preventing infections and gastrointestinal (peptic ulcer, dysphagia, pancreatitis), neurological (e.g. epilepsy, ischemic cerebrovascular disease, stroke), reproductive (infertility, ovulatory defects, endometriosis), autoimmunity/inflammatory (actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease), and cell proliferative disorders including cancers (of the breast, adrenal gland, bone). They may also be used to treat fetal/familial insomnia, nutritional and metabolic diseases of the nervous system, myopathies, mental disorders (anxiety, schizophrenia, mood), as well as infections caused by parasites (malaria, leishmania, trypanosoma), viral (adenovirus, coronavirus, flavivirus), bacterial (e.g. pneumococcus, staphylococcus, bacillus), and fungal (aspergillus, blastomycosis, dermatophytes) agents. The nucleic acids, polypeptides, antagonists, agonists, pharmaceutical compositions, and antibodies may also be used for treating or preventing disorders associated with increased or decreased expression or activity of EXCS. EXCS polynucleotides may also be used to detect and quantify gene expression in biopsied tissues in which expression of EXCS may be correlated with the disease, to determine presence or excess expression of EXCS, to monitor regulation of EXCS levels during therapeutic intervention, to detect the presence of associated disorders, as targets in microarray, to generate hybridization probes, and to detect differences in gene sequences among normal, carrier or affected individuals. Antibodies may also be used in diagnosing disorders, in monitoring patients being treated with EXCS agonists, antagonists or inhibitors. Sequences AAB48057-B48082 represent the EXCS of the invention.

Query Match	Similarity	Score	ID	Length
100.0%;	100.0%;	1243;	DB 22;	342;
Best Local Similarity	100.0%;	Pred. No. 3	3e-102;	
Matches 229;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Query	1	MARHGGCSLTITVPAWAAIIALLALGYERALALPEICTGCGSVQNIISKVAFYCKTTREML	60
Db	114	mahpqpsstltlvpwaaallalgyeralalpelctqpgsvqnlskafyckltrelml	173
QY	61	HARCCCNOKGTTIGLDLONCSLEDPGPNPHOAHNTVIIIDLQANPLKGLIANTFRGFTOLQ	120
Db	174	harcclngqstlglidlgncslcdpgpnfhqhtltvldlgamplkgdlantlrftgclq	233
QY	121	TLLLPDHVNCPPGINAMNTITSYIDNQLCGGKNCNNTGDEPCPCENGSCVPDGPGLQ	180
Db	234	tlllpdhvncppginamntitsyidnqlcgqknlcnntgdcpcpcengscvpdpqjllq	293
QY	181	CVCADCFHGYKCMROGGSFLMFFGLIGATTLSVSIILMANQRRKATS	229
Db	294	cvcadcfhgykcmrgsfslmffgligattlsvsiilwaqrrkats	342

RESULT	10
AA05282	
ID	AA05282 standard; Protein; 229 AA.
XX	
AC	AA05282;
XX	
DT	22-JUN-1999 (first entry)
XX	
DE	EGF-like homologue PRO240.
XX	
KW	Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;
KW	EBAR-2; inhibitor; tumour growth; cancer; EGF-like homologue;
KW	FGF-8 homologue.
XX	
OS	Homo sapiens.
XX	
PN	MO9914327-A2.
XX	
PD	25-MAR-1999.
XX	
FF	10-SEP-1998; 98WO-US18824.
XX	

PR 25-NOV-1997; 97US-0066940.  
 PR 17-SEP-1997; 97US-0059114.  
 PR 17-SEP-1997; 97US-0059117.  
 PR 18-SEP-1997; 97US-0059263.  
 PR 15-OCT-1997; 97US-0062125.  
 PR 17-OCT-1997; 97US-0062285.  
 PR 17-OCT-1997; 97US-0062287.  
 PR 24-OCT-1997; 97US-0062816.  
 PR 29-OCT-1997; 97US-0063704.

XX (GENTH ) GENENTECH INC.

XX Bolstein D, Goddard A, Gurney A, Hillan K, Lawrence DA,  
 PI Roy M, Wood WI;

XX WPI: 1999-229532/19.  
 DR N-PSDB; AAX28432.

XX Antibodies against specific proteins overexpressed in tumours

XX Example 1; Fig 12; 130pp; English.

XX This sequence represents the EGF-like homologue PRO240.  
 CC The invention relates to antibodies (Ab) that bind to any of the  
 CC polypeptides (I) designated PRO187, PRO533, PRO214, PRO240, PRO211,  
 CC PRO230, PRO261, PRO246 or EBAF-2. The Ab, or other agents that inhibit  
 CC expression and/or activity of (I) are used: (1) to inhibit growth of  
 CC tumours; and (1i) as diagnostic/prognostic reagents for detection or  
 CC quantification of (I) in cells or tissues, by standard immunoassays, with  
 CC overexpression being indicative of cancer. For therapeutic use, the Ab  
 CC may be conjugated to a toxin, chemotherapeutic agent or radioisotope.  
 CC Genes expressing (I), many of which are growth factor homologues, are  
 CC overexpressed in some cases of cancer.

XX Sequence 229 AA;

Query Match 99.0%; Score 1231; DB 20; Length 229;

Best Local Similarity 99.1%; Pred. No. 2.4e-101;  
 Matches 227; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAPHGSGSLTLVPMWAAALLALGVERALALPEICTQCPGVSQNLKVAFYCKTTRLEML 60  
 Db 1 maphgsgsltlvpmwaaallalgyeralalpeictqcpvsnlskvaftcttrelml 60  
 QY 61 HARCCCLNOKGTTLLGLDLONCSLEDPGPNFHQAHTTVIIDLOANPLKGDLANTFRGFTQLQ 120  
 Db 61 haccclnkgttllglldqncsleddpgpnfhqahcttviiidlganplkgdlantrfgrftqlq 120  
 QY 121 TLILPOHVNCPGGINAMNTTSYIDNOCOGOKNLNNTGDPKPCPKSGCVPDPGPGLLQ 180  
 Db 121 tlllpqhvnpcpgginamnttsyidnqlcggknlcnntgdpemcpenegscvdpdpqllq 180  
 QY 181 CVCADGFHGYKCMROGFSFLMFGILGATTLVSILLMAQORRKAATS 229  
 Db 181 cvcadgffgykcmrgsfslmffgiltgattlsvsillwatqrkaks 229

RESULT 11

AAG00157 standard; Protein: 147 AA.

XX AAG00157;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 4238.

KW Human, 5' EST, expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX

PM EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GENTH ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-500381/45.

XX N-PSDB; AAC00163.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX Claim 13; SEQ ID 4238; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 147 AA;

Query Match 63.6%; Score 790; DB 21; Length 147;

Best Local Similarity 100.0%; Pred. No. 1.6e-62;  
 Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPHGSGSLTLVPMWAAALLALGVERALALPEICTQCPGVSQNLKVAFYCKTTRLEML 60  
 Db 1 maphgsgsltlvpmwaaallalgyeralalpeictqcpvsnlskvaftcttrelml 60  
 QY 61 HARCCCLNOKGTTLLGLDLONCSLEDPGPNFHQAHTTVIIDLOANPLKGDLANTFRGFTQLQ 120  
 Db 61 haccclnkgttllglldqncsleddpgpnfhqahcttviiidlganplkgdlantrfgrftqlq 120  
 QY 121 TLILPOHVNCPGGINAMNTTSYIDNOCOGOKNLNNTGDPKPCPKSGCVPDPGPGLLQ 147  
 Db 121 tlllpqhvnpcpgginamnttsyidnqlcggknlcnntgdpemcpenegscvdpdpqllq 147

RESULT 12

AAU04296 standard; Protein: 52 AA.

XX AAU04296;

XX 24-OCT-2001 (first entry)

XX Transforming growth factor (TGF) alpha HI.

KW Human, TGF alpha HI; transforming growth factor alpha HI; cancer;  
 KW diagnostic; therapeutic; immune disorder; multiple sclerosis;  
 KW systemic lupus erythematosus; human immunodeficiency virus; HIV;  
 KW hyperproliferative disorder; Gaucher's disease; cardiovascular disease;  
 KW Sicilian syndrome; Chaga's cardiomyopathy; coronary arteriosclerosis;  
 KW angiogenic disorder; corneal graft; neovascularization; wound healing;  
 KW diabetic retinopathy; neurological disorder; Huntington's chorea;  
 KW Alzheimer's disease; Parkinson's disease.

XX Homo sapiens.  
 OS  
 XX  
 XX WO200140251-A1.  
 PN  
 XX  
 XX 07-JUN-2001.  
 PD  
 XX  
 XX 01-DEC-2000; 2000WO-US32745.  
 PF  
 XX  
 XX 02-DEC-1999; 99US-0168387.  
 PR  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX Wei Y;  
 PI  
 XX  
 XX WPI: 2001-441480/47.  
 DR  
 XX  
 XX Nucleic acid encoding human transforming growth factor alpha III  
 PT (TGF $\alpha$ ), useful for preventing, diagnosing and/or treating e.g. Cancer  
 PT and Parkinson's disease -  
 PS  
 XX  
 XX Disclosure; Fig 2; 302pp; English.

CC The sequence represents the amino acid sequence of human transforming  
 CC growth factor (TGF) alpha III. TGF alpha III nucleic acid and protein  
 CC may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate polypeptide expression, for example immune  
 CC disorders (e.g. multiple sclerosis, systemic lupus erythematosus and  
 CC human immunodeficiency virus (HIV) infections), hyperproliferative  
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases  
 CC (e.g. Schmitt's syndrome, Chaga's cardiomyopathy and coronary  
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft  
 CC neovascularisation and diabetic retinopathy), neurological disorders  
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),  
 CC infectious diseases and/or for promoting wound healing, regeneration  
 CC and/or chemotaxis (full details given in specification). Additionally,  
 CC the nucleic acid may be used to produce the secreted polypeptide,  
 CC inserting the nucleic acids into a host cell and culturing the cell to  
 CC assess the protein. It may also be used as a DNA probe in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acid  
 CC sequences in samples, and therefore which patients may be in need of  
 CC restorative therapy. The polypeptides may also be used as antigens in the  
 CC production of antibodies against TGF alpha III and in assays to identify  
 CC modulators of TGF alpha III. The anti-TGF alpha III antibodies may also  
 CC be used as diagnostic agents for detecting the presence of TGF alpha III  
 CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)).  
 CC  
 XX

Sequence 52 AA:

Query Match 24.4%; Score 303; DB 22; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-20;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 151 GOKNICNNTGDEPCNGSCVPDGPGLQCACADGFCYKCMRGSSLLM 202  
 Db 1 gqknlcntgdpemcngscvpdgpjllqcvcadfngykcmrgssllm 52

RESULT 13  
 AAM24243  
 ID AAM24243 standard; Protein: 173 AA.  
 XX  
 XX AAM24243;  
 AC  
 XX  
 XX 12-OCT-2001 (first entry)  
 DT  
 XX  
 XX Rat EST encoded protein seq ID NO: 1768.  
 DE  
 XX  
 XX

Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KW diagnostics; forensic test; gene mapping; genetic disorder;  
 KW biodiversity; gene therapy; nutrition.

XX Rattus norvegicus.  
 OS  
 XX  
 XX WO200154477-A2.  
 PN  
 XX  
 XX 02-AUG-2001.  
 PD  
 XX  
 XX 25-JAN-2001; 2001WO-US02687.  
 PF  
 XX  
 XX 25-JAN-2000; 2000US-0491404.  
 PR  
 XX 17-JUL-2000; 2000US-0617746.  
 PR 03-AUG-2000; 2000US-0631451.  
 PR 15-SEP-2000; 2000US-0663870.  
 PR  
 XX  
 XX (HSE-) HYSEQ INC.  
 PA  
 XX  
 XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
 PI Cao Y, Dermanac RA, Zhang J, Werhman T;  
 PI  
 XX  
 XX WPI: 2001-476164/51.  
 DR  
 XX  
 XX N-Psds; AAH98902.  
 DR  
 XX  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 PT  
 XX  
 XX Claim 20; Page 1162; 1275pp; English.

CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a  
 CC protein of the invention.  
 CC  
 XX

Sequence 173 AA:

Query Match 8.3%; Score 103; DB 22; Length 173;  
 Best Local Similarity 28.4%; Pred. No. 0.13;  
 Matches 40; Conservative 14; Mismatches 65; Indels 22; Gaps 8;  
 Oy 60 LPARCLNCKTITGLDNLGNSLDEPNEFHQATTYIID-LQANLKGDLANTFPGFT 117  
 Db 8 lhlvcll-----clgpgqpyradccshcdlaqccapqgscrcdp-----gweglh 56  
 Oy 118 QLQTLILP--OHVNCPGGINAMNTIT-SYIDNQICGOKNICNNTGDEPCNGSCVPD 174  
 Db 57 cercvmpgpcqgltch--qpwgclchsgwagkrcdkdehlc-lltqsp--cqnqgqcmvd 110  
 Oy 175 GPGILQVCVADGFCYKCMRG 195  
 Db 111 ggggyhcvclpgfygridcark 131

RESULT 14  
 AAE01167  
 ID AAE01167 standard; Protein: 383 AA.  
 XX  
 XX AAE01167;  
 AC  
 XX  
 XX 17-JUL-2001 (first entry)  
 DT  
 XX  
 XX Human gene 4 encoded secreted protein HKAANV61, SEQ ID NO:68.  
 DE  
 XX  
 XX

Human: secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiotensin disorder; kidney disorder;



```

Query Match      8.28; Score 102.5; Db 21; Length 1404;
Best Local Similarity 22.78; Pred.No.1.6;
Matches 40; Conservative 24; Mismatches 65; Indels 47; Gaps 11;

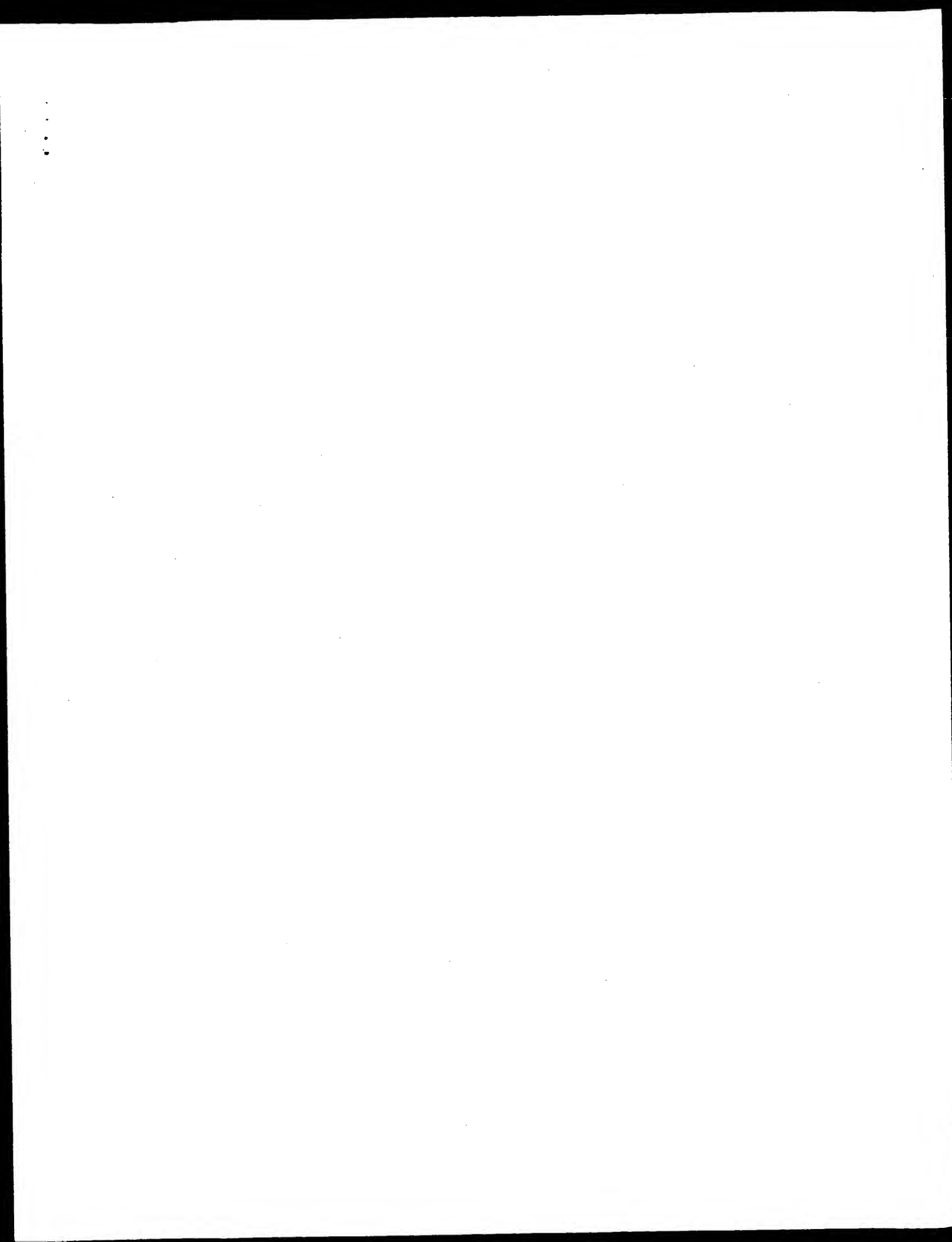
QY 37 QCPGVSVNLKFAVYCKTTRMLHARC-----CLNKGSTI-----LGLDQNS 81
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 qcaflyntctctf-cpriddgfhagcssegkiclmgwgvnceaalckagcpryhk 294
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 82 LEDGGPMHQAHHTVILIQANP-LKGDLANFRFGTLOLILP--QHYNCGEGINAMN 138
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 cdprg-----ecetprgywgrcn-----ecmyvgckhbscng--sawk 332
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 139 TI--TSTIDNQIQGQRKLNCTNTGDEPMKPENGSCVPDPGLTLCYCAADCFHYK 192
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 cvedtnw-ggllcdgdlncf--gtnepockhgtycentapdkyrcytaeaglsqgc 364
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: March 28, 2002, 08:58:40  
Job time: 5167 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:01:41 ; Search time 59.64 Seconds

(without alignments)  
284.420 Million cell updates/sec

Title: US-09-726-348-2

Perfect score: 229

Sequence: 1 MAPHGPSLTLVPMMAALL.....TTLVSILLWATORKAKTS 229

Scoring table: OLIGO

Searched: 522463 seqs, 74073290 residues

Word size: 0

Total number of hits satisfying chosen parameters: 237521

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database: A\_Geneseq.1101.\*

```
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	3.1	23	22	Human transmembran
2	7	3.1	24	22	Maize hydroxyproli
3	6	2.6	9	22	HLA binding TADG-1
4	6	2.6	9	22	HLA binding TADG-1
5	6	2.6	9	22	HLA binding TADG-1
6	6	2.6	9	22	HLA binding TADG-1
7	6	2.6	9	22	HLA binding TADG-1
8	6	2.6	9	22	HLA binding TADG-1
9	6	2.6	9	22	HLA binding TADG-1
10	6	2.6	10	21	C. maltosa cytochr
11	6	2.6	12	15	Protonectarina pep

12	6	2.6	14	22	AA845719
13	6	2.6	15	16	AA876703
14	6	2.6	15	21	AA811403
15	6	2.6	18	18	AA836637
16	6	2.6	18	18	AA809486
17	6	2.6	18	22	AA836724
18	6	2.6	19	21	AA822969
19	6	2.6	19	21	AA822985
20	6	2.6	19	22	AA864214
21	6	2.6	20	20	AA809157
22	6	2.6	20	21	AA822979
23	6	2.6	21	19	AA870528
24	6	2.6	21	21	AA822980
25	6	2.6	21	21	AA822986
26	6	2.6	24	21	AA822988
27	6	2.6	25	19	AA860083
28	6	2.6	25	20	AA814821
29	6	2.6	25	21	AA822987
30	6	2.6	26	20	AA831494
31	6	2.6	26	21	AA825853
32	6	2.6	26	21	AA809081
33	6	2.6	28	22	AA818225
34	6	2.6	28	22	AA830716
35	6	2.6	28	22	AA805838
36	6	2.6	30	20	AA809284
37	5	2.2	6	16	AA880654
38	5	2.2	6	21	AA896772
39	5	2.2	6	22	AA874917
40	5	2.2	7	15	AA856275
41	5	2.2	7	17	AA896829
42	5	2.2	7	19	AA862062
43	5	2.2	7	19	AA853903
44	5	2.2	7	21	AA828277
45	5	2.2	7	22	AA890878

#### ALIGNMENTS

RESULT 1	
AA835198	standard: protein: 23 AA.
AC	AA835198;
XX	
DT	24-APR-2001 (first entry)
XX	
DE	Human transmembrane MUC3 fragment #4.
XX	
KW	Human: trc3-membrane MUC3; chromosome 7q22; colorectal cancer; inflammatory bowel disease; chronic bronchitis; asthma; cystic fibrosis.
KW	
OS	Homo sapiens
XX	
PN	W0200104152-11
XX	
PD	18-JAN-2001
XX	
PF	13-JUL-2000; 2000MO-AU00846.
XX	
PR	13-JUL-1999; 99NZ-0336726.
XX	
PA	(MCGU/) MCGUCKIN M A.
PA	(WILL/) WILLIAMS S J.
XX	
PI	McGuckin MA, Williams SJ.
XX	
DR	WPI: 2001-138317/14.
XX	
PT	Novel transmembrane mucin 3 protein useful for producing
PT	anti-transmembrane mucin 3 antibodies useful in diagnosis and prognosis
PT	of colorectal cancer, inflammatory bowel disease and detecting
PT	transmembrane mucin 3 -

xx Claim 1; Page 32; 42pp; English.

cc The present invention provides the protein and coding sequences of the  
cc human transmembrane protein MUC3. The gene encoding this protein is found  
cc on human chromosome 7q22. The sequences can be used in the diagnosis of,  
cc detection of predisposition to and prognosis of inflammatory bowel  
cc disease, colorectal cancer, chronic bronchitis, asthma, cystic fibrosis  
cc and epithelial cancers, and in the identification of treatments for the  
cc same diseases. The present sequence is a fragment of the MUC3 protein.

xx Sequence 23 AA;

Query Match 3.1%; Score 7; DB 22; Length 23;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 LLLALGV 25  
Db 16 LLLALGV 22

# RESULT 2

AAB66904 standard; Peptide: 24 AA.

xx AAB66904;

xx 12-APR-2001 (first entry)

xx Maize hydroxyproline-rich glycoprotein signal peptide.

xx Insecticide; transgenic plant; insect-resistance; signal peptide.

xx Zea mays.

xx WO200100841-A1.

xx 04-JAN-2001.

xx 23-JUN-2000; 2000WO-GB02457.

xx 29-JUN-1999; 99GB-0015215.

xx 23-DEC-1999; 99GB-0030536.

xx (ZENE) ZENECA LTD.

xx Griffin J, Carlile AJ, Cayley PJ, Mackay EA, Warner SAJ;

xx Vincent JL, Lee MD;

xx WPI: 2001-123015/13.

xx Novel insecticidal protein obtained from species of Paecilomyces for  
xx controlling insects, and for insect-resistant transgenic plant  
xx production

xx Disclosure; Pages 47-48; 72pp; English.

cc The present invention relates to novel insecticidal proteins obtained  
cc from Paecilomyces sp. (see AAB66899 to AAB66901 and AAB66913). The  
cc insecticidal proteins can be used to produce transgenic plants, which  
cc are insect-resistant. Also, the insecticidal proteins are useful for  
cc controlling insects by providing them at a locus where insects feed. The  
cc present sequence is a signal peptide used in the present invention.

xx Sequence 24 AA;

Query Match 3.1%; Score 7; DB 22; Length 24;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 ALLIAL 23  
Db 7 allial 13

# RESULT 3

AU02236 standard; Peptide: 9 AA.

xx AU02236;

xx 29-AUG-2001 (first entry)

xx HLA binding TADG-16 peptide #12.

xx Human; extracellular serine protease; tumour antigen derived gene-16;  
xx TADG-16; ovarian carcinoma; breast cancer; lung cancer; colon cancer;  
xx prostate cancer; HLA type.

xx Homo sapiens.

xx WO200127257-A1.

xx 19-APR-2001.

xx 13-OCT-2000; 2000WO-US28558.

xx 14-OCT-1999; 99US-0418527.

xx (UYAR-) UNIV ARKANSAS.

xx O'Brien TJ, Underwood LJ, Shigemasa K;

xx WPI: 2001-273769/28.

xx New tumour antigen-derived gene-16 protein, useful for diagnosis and  
xx treatment of ovarian, breast, lung, colon and prostate cancer

xx Example 8; Page 52; 124pp; English.

cc AU02225-AU02384 represent TADG-16 peptides which are tested for  
cc their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,  
cc HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour  
cc antigen derived gene-16 protein, TADG-16 (AU02223), is a novel human  
cc extracellular serine protease. TADG-16 is expressed in normal ovaries  
cc and testes and in certain ovarian carcinomas. TADG-16 contains the  
cc conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
cc characteristic of the serine protease family. An antisense  
cc oligonucleotide having a complementary sequence to the TADG-16 nucleic  
cc acid is useful for treating various cancers, including ovarian, breast,  
cc lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
cc antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
cc TADG-16 protein or its fragments are useful for vaccinating an individual  
cc against TADG-16.

xx Sequence 9 AA;

Query Match 2.6%; Score 6; DB 22; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 ALLIAL 23

Db 1 allial 6

# RESULT 4

AU02263 standard; Peptide: 9 AA.

xx AU02263;

xx 14-OCT-1999; 99US-0418527.

```

XX 19-APR-2001.
PD
XX 13-OCT-2000; 2000WO-US28558.
XX
PF 14-OCT-1999; 99US-0418527.
XX
PR (UYAR-) UNIV ARKANSAS.
XX
PA
PI O'Brien TJ, Underwood LJ, Shigemasa K;
XX
DR WPI: 2001-273769/28.
XX
XX New tumour antigen-derived gene-16 protein, useful for diagnosis and
PT treatment of ovarian, breast, lung, colon and prostate cancer -
PT
PS
XX
XX Example 8; Page 54; 124pp; English.
XX
XX AAU02225-AAU02384 represent TADG-16 peptides which are tested for
CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,
CC HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour
CC antigen derived gene-16 protein, TADG-16 (AAU02223), is a novel human
CC extracellular serine protease. TADG-16 is expressed in normal ovaries
CC and testes and in certain ovarian carcinomas. TADG-16 contains the
CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence
CC oligonucleotide having a complementary sequence to the TADG-16 nucleic
CC acid is useful for treating various cancers, including ovarian, breast,
CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and
CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.
CC TADG-16 protein or its fragments are useful for vaccinating an individual
CC against TADG-16.
CC
XX
SQ Sequence 9 AA;

Query Match 2.6%; Score 6; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 ALLAL 23
| | | | |
Db 3 allal 8

RESULT 6
AAU02297
ID AAU02297 standard; Peptide; 9 AA.
XX
AC AAU02297;
XX
DT 29-AUG-2001 (first entry)
XX
DE HLA binding "ADG-16 peptide #73.
XX
KW Human: extracellular serine protease; tumour antigen derived gene-16;
KW TADG-16; ovarian carcinoma; breast cancer; lung cancer; colon cancer;
KW prostate cancer; HLA type.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO200127257-A1.
XX
XX PD 19-APR-2001.
XX
XX PF 13-OCT-2000; 2000WO-US28558.
XX
XX PR 14-OCT-1999; 99US-0418527.
XX
XX PA (UYAR-) UNIV ARKANSAS.
XX
XX PI O'Brien TJ, Underwood LJ, Shigemasa K;
XX
XX
XX O'Brien TJ, Underwood LJ, Shigemasa K;
XX

```

DR WPI; 2001-273769/28.  
 XX New tumour antigen-derived gene-16 protein, useful for diagnosis and  
 PT treatment of ovarian, breast, lung, colon and prostate cancer -  
 XX  
 XX Example 8; Page 54; 124pp; English.  
 PS  
 CC AAU02225-AAU02384 represent TADG-16 peptides which are tested for  
 CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,  
 CC HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour  
 CC extracellular serine protease. TADG-16 (AAU02223), is a novel human  
 CC antigen derived gene-16 protein. TADG-16 is expressed in normal ovaries  
 CC and testes and in certain ovarian carcinomas. TADG-16 contains the  
 CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
 CC characteristic of the serine protease family. An antisense  
 CC oligonucleotide having a complementary sequence to the TADG-16 nucleic  
 CC acid is useful for treating various cancers, including ovarian, breast,  
 CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
 CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
 CC TADG-16 protein or its fragments are useful for vaccinating an individual  
 CC against TADG-16.  
 CC  
 XX Sequence 9 AA:  
 SQ  
 Query Match 2.6%; Score 6; DB 22; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 18 ALLLAL 23  
 Db 2 allal 7  
 RESULT 7  
 ID AAU02315 standard; Peptide: 9 AA.  
 AC AAU02315;  
 DT 29-AUG-2001 (first entry)  
 DE HLA binding TADG-16 peptide #91.  
 XX Human; extracellular serine protease; tumour antigen derived gene-16;  
 KW TADG-16; ovarian carcinoma; breast cancer; lung cancer; colon cancer;  
 KW prostate cancer; HLA type.  
 XX Homo sapiens.  
 OS  
 XX WO200127257-A1.  
 PN 19-APR-2001.  
 PD 13-OCT-2000; 2000WO-US28558.  
 PE 14-OCT-1999; 99US-0418527.  
 PR (UYAR-) UNIV ARKANSAS.  
 PA O'Brien TJ, Underwood LJ, Shigemasa K;  
 PI WPI; 2001-273769/28.  
 DR New tumour antigen-derived gene-16 protein, useful for diagnosis and  
 PT treatment of ovarian, breast, lung, colon and prostate cancer -  
 XX  
 XX Example 8; Page 54; 124pp; English.  
 PS  
 CC AAU02225-AAU02384 represent TADG-16 peptides which are tested for  
 CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,  
 CC HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour  
 CC extracellular serine protease. TADG-16 (AAU02223), is a novel human  
 CC antigen derived gene-16 protein. TADG-16 is expressed in normal ovaries  
 CC and testes and in certain ovarian carcinomas. TADG-16 contains the  
 CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
 CC characteristic of the serine protease family. An antisense  
 CC oligonucleotide having a complementary sequence to the TADG-16 nucleic  
 CC acid is useful for treating various cancers, including ovarian, breast,  
 CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
 CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
 CC TADG-16 protein or its fragments are useful for vaccinating an individual  
 CC against TADG-16.  
 CC

CC extracellular serine protease. TADG-16 is expressed in normal ovaries  
 CC and testes and in certain ovarian carcinomas. TADG-16 contains the  
 CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
 CC characteristic of the serine protease family. An antisense  
 CC oligonucleotide having a complementary sequence to the TADG-16 nucleic  
 CC acid is useful for treating various cancers, including ovarian, breast,  
 CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
 CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
 CC TADG-16 protein or its fragments are useful for vaccinating an individual  
 CC against TADG-16.  
 CC  
 XX Sequence 9 AA:  
 SQ  
 Query Match 2.6%; Score 6; DB 22; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 18 ALLLAL 23  
 Db 2 allal 7  
 RESULT 8  
 ID AAU02334 standard; Peptide: 9 AA.  
 AC AAU02334;  
 DT 29-AUG-2001 (first entry)  
 DE HLA binding TADG-16 peptide #110.  
 XX Human; extracellular serine protease; tumour antigen derived gene-16;  
 KW TADG-16; ovarian carcinoma; breast cancer; lung cancer; colon cancer;  
 KW prostate cancer; HLA type.  
 XX Homo sapiens.  
 OS  
 XX WO200127257-A1.  
 PN 19-APR-2001.  
 PD 13-OCT-2000; 2000WO-US28558.  
 PE 14-OCT-1999; 99US-0418527.  
 PR (UYAR-) UNIV ARKANSAS.  
 PA O'Brien TJ, Underwood LJ, Shigemasa K;  
 PI WPI; 2001-273769/28.  
 DR New tumour antigen-derived gene-16 protein, useful for diagnosis and  
 PT treatment of ovarian, breast, lung, colon and prostate cancer -  
 XX  
 XX Example 8; Page 55; 124pp; English.  
 PS  
 CC AAU02225-AAU02384 represent TADG-16 peptides which are tested for  
 CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,  
 CC HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour  
 CC antigen derived gene-16 protein. TADG-16 (AAU02223), is a novel human  
 CC extracellular serine protease. TADG-16 is expressed in normal ovaries  
 CC and testes and in certain ovarian carcinomas. TADG-16 contains the  
 CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
 CC characteristic of the serine protease family. An antisense  
 CC oligonucleotide having a complementary sequence to the TADG-16 nucleic  
 CC acid is useful for treating various cancers, including ovarian, breast,  
 CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
 CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
 CC TADG-16 protein or its fragments are useful for vaccinating an individual  
 CC against TADG-16.  
 CC

SQ Sequence 9 AA;

Query Match 2.6%; Score 6; DB 22; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 ALLIAL 23  
 |||||  
 Db 2 allial 7

RESULT 9

AAU02351 ID AAU02351 standard; peptide; 9 AA.

XX AC AAU02351;

XX DT 29-AUG-2001 (first entry)

XX DE HLA binding TADG-16 peptide #127.

XX KW Human; extracellular serine protease; tumour antigen derived gene-16;

XX KM TADG-16; ovarian carcinoma; breast cancer; lung cancer; colon cancer;

XX KM prostate cancer; HLA type.

XX OS Homo sapiens.

XX PN WO200127257-A1.

XX PD 19-APR-2001.

XX PF 13-OCT-2000; 2000WO-US28558.

XX PR 14-OCT-1999; 99US-0418527.

XX PA (UYAR-) UNIV ARKANSAS.

XX PI O'Brien TJ, Underwood LJ, Shigemasa K;

XX DR WPI: 2001-273769/28.

XX PT New tumour antigen-derived gene-16 protein, useful for diagnosis and  
 XX PT treatment of ovarian, breast, lung, colon and prostate cancer -  
 XX PS Example 8; Page 55; 124pp; English.

CC AAU02225-AAU02384 represent TADG-16 peptides which are tested for  
 CC their binding affinity to the 8 haplotypes HLA A0201, HLA A0205,  
 CC HLA A1, HLA A24, HLA B7, HLA B8, HLA B2702, and HLA B4403. Tumour  
 CC antigen derived gene-16 protein, TADG-16 (AAU02223), is a novel human  
 CC extracellular serine protease. TADG-16 (AAU02223) contains the  
 CC and testes and in certain ovarian carcinomas. TADG-16 contains the  
 CC conserved catalytic triad, His-Asp-Ser, and a signal secretion sequence  
 CC characteristic of the serine protease family. An antisense  
 CC oligonucleotide having a complementary sequence to the TADG-16 nucleic  
 CC acid is useful for treating various cancers, including ovarian, breast,  
 CC lung, colon and prostate. The TADG-16 nucleic acid, TADG-16 protein and  
 CC antibodies specific to TADG-16 are useful for the diagnosis of cancer.  
 CC TADG-16 protein or its fragments are useful for vaccinating an individual  
 CC against TADG-16.

SQ Sequence 9 AA;

Query Match

Best Local Similarity 2.6%; Score 6; DB 22; Length 9;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 ALLIAL 23  
 |||||  
 Db 4 allial 9

RESULT 10

AAB11404 ID AAB11404 standard; peptide; 10 AA.

XX AC AAB11404;

XX DT 22-FEB-2001 (first entry)

XX DE C. maltosa cytochrome b5 peptide fragment T54.

XX KM Cytochrome b5; alkane metabolism; oxidation; long-chain alkyl compound;

XX KM long-chain dicarboxylic acid.

XX OS Candida maltosa.

XX PN WO200065061-A2.

XX PD 02-NOV-2000.

XX PF 18-APR-2000; 2000WO-DE01246.

XX PR 24-APR-1999; 99DE-1018763.

XX PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

XX PI Schunck W, Chernogolov A;

XX DR WPI: 2000-679674/66.

XX PT Nucleic acid sequences from alkane metabolizing Candida yeast, encoding

XX PT cytochrome b5 and used for the oxidation of long chain alkyl compounds

XX PT and for the production of long chain dicarboxylic acids -

XX PS Example 1.2; Page 6; 27pp; German.

XX CC This invention describes novel nucleic acid sequences from alkane

XX CC metabolizing Candida yeasts, encoding cytochrome b5-polypeptide, its

XX CC fragments, variants and mutations. The nucleic acids and polypeptides

XX CC are used for the oxidation of long-chain alkyl compounds with at least

XX CC 10C, and for the production of long-chain dicarboxylic acids by oxidizing

XX CC n-alkanes and fatty acids with at least 10C.

XX SQ Sequence 10 AA;

OY 140 ITSYTD 145

XX DB 4 itsytd 9

RESULT 11  
 AAR45452 ID AAR45452 standard; peptide; 12 AA.

XX AC AAR45452;

XX DT 11-JUL-1994 (first entry)

XX DE Protonecarina peptide P1 with histamine releasing activity.

XX KM Hunting wasp; poison sac; histamine release; secretion; allergy;

XX KM inflammation; mast cell.

XX OS Protonecarina sylvestrae.

XX FT Key Location/Qualifiers  
 FT Modified-site 12 /note= "in amide form"

XX JP05331194-A.  
 XX 14-DEC-1993.  
 PD 03-JUN-1992; 92JP-0142299.  
 XX 03-JUN-1992; 92JP-0142299.  
 XX 03-JUN-1992; 92JP-0142299.  
 XX (DAIL ) DAICEL CHEM IND LTD.  
 XX WPI, 1994-022916/03.  
 DR BIoactive peptide obtd. from hunting wasp - used for development  
 XX of research reagents and drugs for allergy and inflammation  
 PT related to mast cell  
 XX  
 PS Claim 1; Page 8; 10pp; Japanese.  
 CC Proconectarina peptide P-1 was isolated from the pollen sac of the  
 CC hunting wasp P. sylvestre. The peptide has a high histamine releasing  
 CC activity on rat abdominal mast cells and will be useful for the  
 CC development of research reagents and drugs for allergy and  
 CC inflammation related to mast cells. See also AAR45453 and AAR45454.  
 XX  
 SO Sequence 12 AA:

Query Match 2.6%; Score 6; DB 15; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 GTILGL 75  
 |||||  
 Db 3 gtilgl 8

RESULT 12  
 ID AAB45719  
 AAB45719 standard; Protein; 14 AA.

AC AAB45719;  
 XX  
 DT 15-MAR-2001 (first entry)  
 XX

DE Human 7TM clone H7TME50 protein fragment #6.

XX Transmembrane receptor protein; 7TM; gene therapy; human; bactericidal;  
 KW fungicidal; viricidal; anti-HIV; analgesic; cytostatic; anabolic;  
 KW anti-asthmatic; anti-parkinsonian; cardiac; hypertensive; vaccine;  
 KW osteopathic; anti-ulcer; immunosuppressive; cerebroprotective; infection;  
 KW antidepressant; neuroprotective; hepatolical receptor; infection;  
 KW serpentine receptor; G-protein coupled receptor; cancer; anorexia;  
 KW bulimia; asthma; Parkinson's disease; acute heart failure; hypertension;  
 KW hypertension; urinary retention; osteoporosis; angina pectoris; ulcer;  
 KW myocardial infarction; allergy; benign prostatic hypertrophy; anxiety;  
 KW Schizophrenia; manic depression; Huntington's chorea.

XX Homo sapiens.

XX WO200071584-A1.

XX 30-NOV-2000.

XX 19-MAY-2000; 2000WO-US13737.

XX 20-MAY-1999; 99US-0135167.

XX 13-JUN-1999; 99US-0143616.

XX 09-SEP-1999; 99US-0152934.

XX 14-MAR-2000; 2000US-0189029.  
 XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Ni J, Soppet DR, Li Y, Fan P;  
 XX WPI: 2001-025139/03.  
 DR N-PSDB; AAC82669.  
 XX  
 XX Nucleic acids encoding human 7 transmembrane receptor polypeptides,  
 PT useful for preventing, diagnosing and treating e.g. asthma, Parkinson's  
 PT disease and ulcers -  
 XX  
 XX Claim 1b; Page 282; 288pp; English.

CC This invention describes novel nucleic acid molecules (I) encoding human  
 CC 7 transmembrane (7TM) receptor polypeptides (also called heptahelical,  
 CC serpentine or G-protein coupled receptors). The products of the invention  
 CC have bactericidal, fungicidal, viricidal, anti-HIV, analgesic,  
 CC cytostatic, anabolic, anti-asthmatic, anti-parkinsonian, cardiac,  
 CC hypotensive, hypertensive, osteopathic, anti-ulcer, immunosuppressive,  
 CC cerebroprotective, antidepressant and neuroprotective activity and can  
 CC be used in gene therapy or in a vaccine. (I) and the 7TM receptor protein  
 CC it encodes may be used in the prevention, treatment and diagnosis of  
 CC diseases associated with inappropriate 7TM receptor expression. (I) and  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate  
 CC the presence of similar nucleic acid sequences (PCR) in samples, and hence  
 CC which patients may be in need of restorative therapy. Diseases associated  
 CC with inappropriate 7TM receptor polypeptide expression include for  
 CC example bacterial, fungal and viral (especially human immuno-deficiency  
 CC virus) infections, pain, cancers, anorexia, bulimia, asthma, Parkinson's  
 CC disease, acute heart failure, hypotension, hypertension, urinary  
 CC retention, osteoporosis, angina pectoris, myocardial infarction, ulcers,  
 CC allergies, benign prostatic hypertrophy, anxiety, schizophrenia, manic  
 CC depression and Huntington's chorea.

XX Sequence 14 AA:

Query Match 2.6%; Score 6; DB 22; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 204 FGILGA 209  
 |||||  
 Db 5 fgilga 10

RESULT 13  
 ID AAR76703  
 AAR76703 standard; peptide; 15 AA.

AC AAR76703;  
 XX  
 DT 31-JAN-1996 (first entry)  
 XX

DE Dendritic branched signal peptide, DBP-3.

XX Dendritic; branched; membrane; signal peptide; cerebral; hormones;  
 KW polyllysine core; drug delivery.

XX Synthetic.

XX Key Location/Qualifiers

XX Modified-site 5

XX /label= branch.site

XX /note= "omega amino group in the sidechain of

XX Lys(5) forms a peptide bond with the

XX C-terminus of the tetrapeptide

XX Arg-Leu-Leu-Leu"  
 XX Modified-site 6  
 XX /label= branch.site  
 XX /note= "omega amino group in the sidechain of  
 XX Lys(6) forms a peptide bond with the  
 XX C-terminus of the peptide  
 XX Arg-Leu-Leu-Leu-Lys. Omega amino group

WPI: 2000-679674/66

diagnostically, e.g. to investigate the mechanism of thrombopoietin

Proliferation and growth of thrombopoietin dependent cell lines.

XX  
SQ Sequence 18 AA:

Query Match 2.6%; Score 6; DB 18; Length 18;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 GSFSLL 201  
DB 12 gsfsll 17

Search completed: March 28, 2002, 09:15:53  
Job time: 852 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 01:48:12 ; Search time 1493.67 Seconds

(without alignments)  
7587.728 Million cell updates/sec

Title: US-09-726-348-1\_COPY\_5\_691

Perfect score: 687

Sequence: 1 atgagcctcaccgcccggg.....gccgaaagccaaacttca 687

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 824589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_hgo\_hum:\*  
31: em\_hgo\_inv:\*  
32: em\_hgo\_rod:\*  
33: em\_hgt\_hum:\*  
34: em\_hgt\_inv:\*  
35: em\_hgt\_rod:\*  
36: em\_hgt\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	687	100.0	922	9	BC002846	BC002846 Homo sapi
2	687	100.0	932	6	AX055686	AX055686 Sequence
3	687	100.0	932	6	AX076904	AX076904 Sequence
4	687	100.0	932	6	AX077021	AX077021 Sequence
5	687	100.0	940	9	BC011006	BC011006 Homo sapi
6	687	100.0	1265	6	AX048198	AX048198 Sequence
7	671.4	97.7	915	9	AF275744	AF275744 Homo sapi
8	665.8	82.4	1074	9	AF144055	AF144055 Homo sapi
9	130.6	19.0	162580	2	AC013403	AC013403 Homo sapi
10	130.6	19.0	192554	2	AC013413	AC013413 Homo sapi
11	103.4	15.1	15012	9	AF116241	AF116241 Homo sapi
12	103.2	15.0	123280	9	AC025662	AC025662 Homo sapi
13	96.4	14.0	2056	9	AF015947	AF015947 Homo sapi
14	92	13.4	208448	2	AC084883	AC084883 Mus muscu
15	87.2	12.7	1671	10	HAMCAD	M31621 Syrian hams
16	77.8	11.3	6558	10	AF189010	AF189010 Rattus no
17	76	11.1	55932	2	AC025791	AC025791 Homo sapi
18	70	7.3	50	6	AX055746	AX055746 Sequence
19	50	7.3	50	6	AX076908	AX076908 Sequence
20	50	7.3	50	6	AX077046	AX077046 Sequence
21	40.2	5.9	106264	2	AC084809	AC084809 Homo sapi
22	40.2	5.9	156461	2	AC090555	AC090555 Homo sapi
23	37	5.4	178783	2	AC079486	AC079486 Mus muscu
24	36.8	5.4	222165	2	AL353801	AL353801 Homo sapi
25	36	5.2	152541	2	AC034118	AC034118 Homo sapi
26	36	5.2	181623	2	AC041030	AC041030 Homo sapi
27	35.8	5.2	146302	2	AC073190	AC073190 Homo sapi
28	35.8	5.2	175226	2	AC018859	AC018859 Homo sapi
29	35.6	5.2	159272	9	HSDJ686C3	AL049712 Human DNA
30	35.4	5.2	10706	9	AL354703	AL354703 Human DNA
31	35.2	5.1	70727	9	AC023028	AC023028 Homo sapi
32	35	5.1	182917	9	AL139383	AL139383 Human DNA
33	35	5.1	207567	2	AL161719	AL161719 Homo sapi
34	34.8	5.1	128715	2	AC092891	AC092891 Homo sapi
35	34.8	5.1	134802	2	AL356852	AL356852 Homo sapi
36	34.8	5.1	138888	9	HS800R24	Z99758 Human DNA s
37	34.8	5.1	169733	2	AC068046	AC068046 Homo sapi
38	34.6	5.0	2424	5	LSEB4BUTO	X51410 LaIicauda s
39	34.6	5.0	87074	9	AL451045	AL451045 Human DNA
40	34.6	5.0	87637	9	AL390960	AL390960 Human DNA
41	34.6	5.0	141347	2	AC083820	AC083820 Rattus no
42	34.6	5.0	151328	2	AL136079	AL136079 Homo sapi
43	34.4	5.0	182905	2	AC022885	AC022885 Homo sapi
44	34.4	5.0	134187	9	AC004704	AC004704 Homo sapi
45	34.4	5.0	208924	9	AC018735	AC018735 Homo sapi

#### ALIGNMENTS

RESULT 1  
BC002846 922 bp mRNA PRI 12-JUL-2001  
LOCUS Homo sapiens, Similar to apoptosis related protein APR-3 clone  
DEFINITION IMAGE:3635746, mRNA.

ACCESSION BC002846  
VERSION BC002846.1 GI:13937653  
KEYWORDS  
SOURCE  
ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
Strausberg, R.  
REFERENCE  
1 (bases 1 to 922)  
TITLE Direct Submission  
JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>





Qy 241 tctctgagagaccctgctgcaacttcatcaggacacatacactgtcatcatagacctg 300  
 |||||  
 Db 252 TCTCTGGAGGACCTGCTGCTCAACTTTCATCAGGACACTTACACTGTCTATATAGACTG 311  
 |||||  
 Qy 301 caagcaaacccctcaagaagtgaacttggccaacaccttcggtgacttactagctccag 360  
 |||||  
 Db 312 CAAGCAAAACCCCTCAAGAGTGAAGTGGCCAAACACTCTCCGTGCTTACTACACTCCAG 371  
 |||||  
 Qy 361 actctgatactgcccacacatgtaactgtctctggaggaataatgctctggaatactc 420  
 |||||  
 Db 372 ACTCTGATAGTCCACCAACATGTCACACTGTCTGAGGAATTAATGCTGGAAATACATAC 431  
 |||||  
 Qy 421 acccttatagacaacacacatctgtcaagggaacacacacacacacacacacacacac 480  
 |||||  
 Db 432 ACCTCTTATATAGAACCAATCTGTCAAGGCAAAAGAACCTTTGCAATACACTGGG 491  
 |||||  
 Qy 481 gaccagaagaatgtgctcctgagaatgatctgttaccctgagtcagagctcttgcag 540  
 |||||  
 Db 492 GACCCAGAAATGTCTCCGAGAAATGATCTGTACTGATGCTCCAGGCTTTTGCAG 551  
 |||||  
 Qy 541 tgtgttctgctgagtggttccacagatgataagtgatagcggagcgtctgctcactg 600  
 |||||  
 Db 552 TGTGTTTGTCTGATGGTTTCCATGAGTACAGTATAGCCGACGAGGCTGCTTCTCACTG 611  
 |||||  
 Qy 601 ctatgtctctcggagatctgagagacacacacacacacacacacacacacacacacac 660  
 |||||  
 Db 612 CTATATGTTCTTCGGGATTCCTGGAGCCACACCTATACCTGCTCCTCATTTGCTTGGGG 671  
 |||||  
 Qy 661 acccagcgcgaaagccaagacttca 687  
 |||||  
 Db 672 ACCCAGCGCCGAAAGCCCAAGACTTCA 698  
 |||||

## RESULT 5

LOCUS BC011006 940 bp mRNA PRI 30-JUL-2001  
 DEFINITION Homo sapiens, clone MGC:13322 IMAGE:4103408, mRNA, complete cds.  
 ACCESSION BC011006  
 VERSION BC011006.1 GI:15029605  
 KEYWORDS MGC.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 940)  
 Strausberg, R.  
 Direct Submission  
 Submitted (25-JUL-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: ggaabs-remail.nih.gov

## REFERENCE

AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JUL-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

## REMARK

COMMENT  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: ggaabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Institute for Systems Biology  
 http://www.systemsbio.org  
 contact: amadan@systemsbiology.org  
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia  
 Greene, Mark Keltman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNLN at: http://lmln.gov  
 Series: IRAL Plate: 19 Row: f Column: 23  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 7706360.  
 location/Qualifiers

## FEATURES

source  
 1..940  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="MGC:13322 IMAGE:4103408"

/tissue\_type="Brain, glioblastoma"  
 /clone\_id="NHL\_MGC\_57"  
 /lab\_host="DH10B"  
 /note="Vector: pDNR-LIB"  
 13..702  
 /codon\_start=1  
 /product="Unknown (protein for MGC:13322)"  
 /protein\_id="AAH11006.1"  
 /db\_xref="GI:15029606"  
 /translation="MAPPGSLTTLVPMALLALGVERALALPEITCQPSVON  
 LSKVAYCKTTRRLMLHARCCINQKGLTGLDQNGSLDEPNEHQAHVITVILDA  
 NPLKGLDANFRGFTQLTLLIPQVNCPCGINSINAMITTSYIDNOGOKNLCNNG  
 PEMGPENGSCVDPDGPGLQCVACADGFHYCYCMRQSPSLMFGILGATLVSILL  
 MATORRKRKTS"  
 BASE COUNT 248 a 242 c 215 g 235 t  
 ORIGIN

Query Match 100.0%; Score 687; DB 9; Length 940;  
 Best Local Similarity 100.0%; Pred. No. 6, 4e-202; Mismatches 0; Gaps 0;  
 Matches 687; Conservative 0; Mismatches 0; Indels 0;

Qy 1 atgagcctcagagcccggtgagcttcaagaccctgtgtcctgtggtcgtccgtctc 60  
 |||||  
 Db 13 ATGGCGCTCACGGGCGCGGCTAGTCTTACGACCTGTGCTCCCTGGGCTGCCCTGCTC 72  
 |||||  
 Qy 61 ctgcctctggtgtgaaaggtctctggtcctaccggagatctgacccaatgtccagg 120  
 |||||  
 Db 73 CTCGCTCTGGGGGTGAAAGGGCTCTGGGCTACCCGAGATATGCAACCAATGTCCAGGG 132  
 |||||  
 Qy 121 agcgtcaaatctgtcaaaagtgtccttcttcttaaaagacacagagctaaagtctg 180  
 |||||  
 Db 133 AGCGTCGCAAAATTTGTCAAAAGTGGCCTTTATTTGAATACACACAGAGTATAGCTG 192  
 |||||  
 Qy 181 catgcccgttgcctcgtgaatcagaagggacacacacacacacacacacacacacacac 240  
 |||||  
 Db 193 CATGCCCGTGTGCTGCTGATGATGAGAGGACACCATCTGGGGCTGATCTCCAGAACTGT 252  
 |||||  
 Qy 241 tctctgagagaccctgtgtcacaacttcaagagacacacacacacacacacacacacac 300  
 |||||  
 Db 253 TCTCTGAGGACCCCTGCTCAAACTTTCTATCAGGACATACCTGATCATCATAGACTG 312  
 |||||  
 Qy 301 caagcaaacccctcaagaagtgaacttggccaacacacacacacacacacacacacacac 360  
 |||||  
 Db 313 CAAGCAAAACCCCTCAAGAGTGAAGTGGCCAAACACTTCCGGCTTATACAGCTCCAG 372  
 |||||  
 Qy 361 actctgatactgcccac 420  
 |||||  
 Db 373 ACTCTGATAGTCCCAACATGTCACACTGTCTGAGGAAATTAATGCTGGAATATATATC 432  
 |||||  
 Qy 421 acccttatagacaac 480  
 |||||  
 Db 433 ACCTCTTATATAGAACCAATCTGTCAAGGCAAAAGAACCTTTGCAATACACTGGG 492  
 |||||  
 Qy 481 gaccagaagaatgtgctcctgagaatgatctgttaccctgagtcagagctcttgcag 540  
 |||||  
 Db 493 GACCCAGAAATGTCTCCGAGAAATGATCTGTACTGATGCTCCAGGCTTTTGCAG 552  
 |||||  
 Qy 541 tgtgttctgctgagatgttccatagatacaagtgatagcgccagggctgtctcactg 600  
 |||||  
 Db 553 TGTGTTTGTCTGATGGTTTCCATGAGTATCAAGTATAGCCCAAGGCTCTTCTCATGTG 612  
 |||||  
 Qy 601 ctatgtctcctgagatctgagagacacacacacacacacacacacacacacacacac 660  
 |||||  
 Db 613 CTATGTTTCTGGGATTCCTGGAGCCACACTATACCTGCTCATTTGCTTGGGGG 672  
 |||||  
 Qy 661 acccagcgcgaaagccaagacttca 687  
 |||||  
 Db 673 ACCCAGCGCCGAAAGCCCAAGACTTCA 699  
 |||||

RESULT 6  
 AX048198

```

LOCUS      AX048198      1265 bp      DNA
DEFINITION Sequence 38 from Patent WO0070049.
ACCESSION  AX048198
VERSION     AX048198.1  GI:11876988
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1265)
AUTHORS     Tang,Y.T., Yue,H., Lal,P., Burford,N., Bandman,O., Baughn,M.R.,
            Azimzai,Y., Lu,D.A. and Paterson,C.
TITLE       Extracellular signaling molecules
JOURNAL     Patent: WO 0070049-A 38 23-NOV-2000;
            Incyte Genomics, Inc. (US)
FEATURES
SOURCE      1..1265
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /note="Incyte ID No: 2207183CH1"

BASE COUNT  301 a      341 c      356 g      267 t

ORIGIN
Query Match      100.0%; Score 687; DB 6; Length 1265;
Best Local Similarity 100.0%; Pred. No. 6, 5e-202;
Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 atggcgctcagcgccgggtagcttctacgacctgtgcccgtggtccgcgcctgctc 60
DB      355 ATGGCGCTCAGCGGCCGGGTAGCTTACGACCTGGGCTGGCCCTGCGCCCTGCTC 414

QY      61 ctgcctcggcggtggaagggctctggtgctacccgagatgacccaatgtccagg 120
DB      415 CTCGCTCGGGCGTGGAAGGGCTCTGGCGCTACCCGAGATATCACCCATGTGCGAGG 474

QY      121 agcgtgcaaatctgtcaaaagtgccttattgtlaaaacgacagagactaagtctg 180
DB      475 AGCGTGCAAAATTTGTCAAAAGTGCCCTTTATTGTAAACAGACGAGAGCTAATCTG 534

QY      181 catgcccgtgtgctgtaaatcaagaaggacacatcttggtgctgtaattccagaactgt 240
DB      535 CATGCCCTTCTGCTGTAATCAAGAAGGGCACCATCTTGGGGCTGGATCTCCAGAACTGT 594

QY      241 tctctggagagacctgtgtccaaacttcatagggacacatcactgtcatatagacctg 300
DB      595 TCTCTGGAGAACCTGTGTCAAAATTTTCATCAGGCACATCCTGTCATCATATGACCTG 654

QY      301 caagcaaacccccctcaaaagtgaacttggccaacacctcgttgcttactcaagctccag 360
DB      655 CAAGCAAAACCCCTCAAAAGTGACTGTGGCCAAACACCTTCCGTGGCTTACTCAGCTCCAG 714

QY      361 actctgatactgccaacaatgtcaactgtctcgtggaagaaatgaatgcttgaatactatc 420
DB      715 ACTCTGATACAGCCCAACATGTCAATGTCCTGAGAGAAATTAATGCGTGAATACATATC 774

QY      421 acccttctatagaacaacaatctgtcaagggtgcaaaagaccttggataaacacaggg 480
DB      775 ACCTTTATATAGACACCAACATCTGTCAAGGGCAAAAGACCTTTGCAATACACAGTGG 834

QY      481 gaccagaagaatgtgtcctgagaatgtaactgtgtacactgtgtccaggtccttgcag 540
DB      835 GACCCAGAAAGTGTCCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 894

QY      541 tgtgttctgctgatagtgttccatagataagaagtgtatgagccaggggtgttctactg 600
DB      895 TGTGTTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 954

QY      601 ctatgtctctcgagatctctggaagccacactatcagctcactatccttcttgggag 660
DB      955 CTTATGTTCTTGGGATCTCTGGAGACCAACATCTATCCGCTCCATCTCTGCTTTGGGCG 1014

QY      661 acccagcgccggaagaagccaagacttca 687

```

```

|||||
DB      1015 ACCCAGCGCCGGAAGCCAGACTTCA 1041

RESULT      7
LOCUS      AF275744
DEFINITION Homo sapiens p18 mRNA (p18) mRNA, complete cds.
ACCESSION  AF275744
VERSION     AF275744.1  GI:14582226
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 915)
AUTHORS     Yang,Y.C., Chen,S.Y. and Chang,M.S.
TITLE       Cloning and characterization of p18
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 915)
AUTHORS     Yang,Y.C., Chen,S.Y. and Chang,M.S.
TITLE       Direct Submission
JOURNAL     Submitted (06-JUN-2000) Dept. of Medical Research, Mackey Memorial
            Hospital, 45 Minshen Rd., Tamsui, Taipei 251, Taiwan
            Location/Qualifiers
FEATURES
SOURCE      1..915
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /gene="p18"
            /gene="p18"
            /gene="p18"
            /product="p18 protein"
            /protein_id="AAK69412.1"
            /translation="MLHARCCCLNOKGTTILGLDNLONCSLEDPENFHOAHITVIIDLA
            NPLKGLDNLANTFERTOLQTLIPQVHNPQGINAWNTITSYIDNOCQGNLNNNG
            DEMCPENSCVDPDGPGLIGVCVAGFGRKMGQSFSLMFGLIATVLSVILL
            MAORRRKAKTS"

BASE COUNT  223 a      244 c      213 g      235 t

ORIGIN
Query Match      97.7%; Score 671.4; DB 9; Length 915;
Best Local Similarity 99.3%; Pred. No. 4, 4e-197;
Matches 686; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

QY      1 atggcgctcagcgccgggtagcttctacgacctgtgcccgtggtccgcgcctgctc 60
DB      9 ATGGCGCTCAGCGGCCGGGTAGCTTACGACCTGGGCTGGCCCTGCGCCCTGCTC 68

QY      61 ctgcctcggcggtggaagggctctggtgctacccg----agatagaccatgtgtcc 116
DB      69 CTCGCTCGGGCGTGGAAGGGCTCTGGCGCTACCCGAGATATCACCCATGTGCGAGG 128

QY      117 agggaggttcaaaatctgtcaaaagtgtgaccttlaatgtlaaaacgacagagactaat 176
DB      129 AGGGAGGCTCAAAATTTGTCAAAAGTGCGCTTTATTTGAACACACAGAGACTAAT 188

QY      177 gctgcatgccgctgtgctgctgaatcaagaaggacacatcttggggctgtgattccagaa 236
DB      189 GCTGCAATGCCCGTGTGCTGCTGTAATCAAGAAGGGCACACATCTTGGGGCTGGATCTCCAGAA 248

QY      237 ctgttctctggaagacctgtgtccaaacttcaatgaaggacacatcactgtatcaataga 296
DB      249 CTGTTCTCTGGAAGACCTTGTCCAAACTTCAATCAGGCACATACCATGTCATCATTA 308

QY      297 ccttgaagcaaacccctcaaaagtgtactgtgccaacaccttccgtgtgttactaagct 356
DB      309 CTTGCAAGCAACCCCTCAAAAGGTGACTTGGCCACACACTTCCGTGCTTACTCAGCT 368

QY      357 ccagactctgatactgtccacaacaatgtcaactgttctctggaagaaatgaatgcttgaatac 416

```

```

|||||
Db 369 CCAGACTCTGATACAGCCACACATGTACTCTCTGGAGGAATTAATGCTGGAAATAC 428
QY 417 taccactcttatagacacacaaatctgtcaaggagcaaaagaccccttgcaataaac 476
Db 429 TATACCTCTTATATAGAACCAACCAATCTGTCAAGGGCAAAAGACCTTGTGCAATACAC 488
QY 477 tggggaaccagcaaatgtgtccctgagaatgagatctgttacttgatgttcgaaggtcttt 536
Db 489 TGGGGAACCAAGAAATGTGCTGCAAGATGATCTTGTACTGATGGTCCAGGTCTTTT 548
QY 537 gcaagtgtgttgctgcatgagtttccatgagatacaagtgtatgcgcaaggctcgtctc 596
Db 549 GCAGTGTGTTGTGCTGATGAGTTTCCATGATACAAAGTATGCGCCAGGGGTCTTCTC 608
QY 597 actgcttatgtctctcggagatctcggagacacacactatcgtctcattcgtcttg 656
Db 609 ACTGCTTATGTTCTCTGGGATCTCTGGAGCCACACTCTATCCGCTCCATCTGCTTGTG 668
QY 657 ggcgacccagcgccgaaagacgaacttca 687
Db 669 GCGGCGCCAGCGCCGCAAAAGCCAAAGCTTCA 699

RESULT 8
AF144055 1074 bp mRNA PRI 15-JUN-1999
LOCUS Homo sapiens apoptosis related protein APR-3 mRNA, complete cds.
ACCESSION AF144055
VERSION AF144055.2 GI:4982484
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Zhu, F., Yan, W., Chal, Y. B., Shao, C., Peng, W. D., Yang, A. G., Wang, C. J.
and Zhao, Z. L.
JOURNAL Direct Submission
Submitted (20-APR-1999) Dept. Biochemistry and Molecular Biology,
Forth military Medical University, No. 17 Changle West Rd., Xi'an,
Shaanxi 710032, P. R. China
2 (bases 1 to 1074)
Zhu, F., Yan, W., Chal, Y. B., Shao, C., Peng, W. D., Yang, A. G., Wang, C. J.
and Zhao, Z. L.
JOURNAL Direct Submission
Submitted (03-JUN-1999) Dept. Biochemistry and Molecular Biology,
Forth military Medical University, No. 17 Changle West Rd., Xi'an,
Shaanxi 710032, P. R. China
Sequence update by submitter
On Jun 3, 1999 this sequence version replaced gi:4868438.
REMARK
COMMENT Location/Qualifiers
FEATURES
SOURCE
1..1074
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="apoptotic HU-60"
/note="cells induced by retinoic acid"
251..877
/codon_start=1
/product="apoptosis related protein APR-3"
/protein_id="A4D3137.2"
/db_xref="GI:4982485"
/translation="MHMSREKAFYKSGLYCKTTRIMHARCCLNKGCTILGLDQ
NCSLEDPGRFHQAHTVITIDLANPLKGLDPLATFRFTDLOTLILPQHYNCGGGINA
WNTTSTIDNODICGKNNNTGDEPMPENSCVDPDGLQVCVCAQCFHGIQVYA
PGLVLAIVYVIRLSCSHSIRLSALGSPARKSODPMYVIGLTIDRSI"
BASE COUNT 262 a 292 c 257 g 263 t
ORIGIN

```

Query Match 82.4%; Score 565.8; DB 9; Length 1074;  
 Best Local Similarity 99.3%; Pred. No. 2.6e-164;  
 Matches 589; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

```

QY 96 cgaatatgcaccccaatgtccaggagcgctgcacaaattgttcaaaagtgtcccttattg 155
Db 245 CCAATATGACACCAATGTCAGGAGACCTGCAAAATTTGTCAAAAGTGACC-TTTATTG 303
QY 156 taaacagacagagagctaaatgtcgaatgcgcgttctgtcctbaatccaaagggacacat 215
Db 304 TAAACAGACAGAGAGTAAATGCTGCATGCCCTTGTGCTTAATCAGAAAGGACCAAT 363
QY 216 ctggggctgagatcccaagactgttctctggaagacccctgtgccaaacttcaatgaagc 275
Db 364 CTGGGGCTGAGATCTCCAGAACTGTTCTGTGAGAGACCTGTGCAAACTTCATCATCAGGC 423
QY 276 acataaccagctatcatatgaactgtcaagaagaacccccctcaagtgacttgccaacac 335
Db 424 ACATACCACTGTATCATATGACCTGCAAGCAAAACCCCTCAAAAGTGACTTGCCCAAC 483
QY 336 ctcccggtgttactacagctccagactgtatcactgccaacaacatgtcaactgtctgg 395
Db 484 CTTCGGTGGCTTTACTAGCTCCAGACTGTGATACGCCACACATGTCAACTGTCTCTGG 543
QY 396 aggaatgaatgctgtgaatactatcaactctatataagacaacaaatctgtcaaggca 455
Db 544 AGGAATTAATGCCCTGGAACTACTATCACTCTTATATAGAACCAACCAATCTGTCAAGGCA 603
QY 456 aaagaaccttgcaataacacctggggagccagaaatgtlccctggaatgagatctgtgt 515
Db 604 AAAGAACCTTTGGCAATACACTGGGAGCCGAAAGTGTCTCGAAGATGATCTTGTGT 663
QY 516 acctgatgtcagagctcttgagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 574
Db 664 ACCTGATGTCACAGGCTTTTGTGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 723
QY 575 gtatggcagagctcgtgtctcactgtatgtcttctggagatctcgtggagccacacac 634
Db 724 GTATGGCCAGGCGCTGCTCTACTGCTATGTTCTTGGGATTTGCGAGACACACATC 783
QY 635 taccgcttcctatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 687
Db 784 TATCGCTTCTCATTTGCTTTGGCGGCCGACGCGCAAAAGCCAAAGCAAGCTTCA 836

RESULT 9
AC013403/c
LOCUS AC013403 162580 bp DNA HTG 22-MAR-2001
DEFINITION Homo sapiens chromosome UNK clone RP11-195B17, WORKING DRAFT
ACCESSION AC013403
VERSION AC013403.8 GI:13431116
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Waterston, R. H.
JOURNAL Unpublished
2 (bases 1 to 162580)
TITLE The sequence of Homo sapiens clone
AUTHORS Waterston, R. H.
JOURNAL Direct Submission
Submitted (09-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 22, 2001 this sequence version replaced gi:11878509.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_NH0195B17
----- Summary Statistics -----

```

Sequencing vector: M3; 48%  
Sequencing vector: pLM3; 37%  
Chemistry: Dye-terminator EM; 48% of reads  
Chemistry: Dye-terminator Big Dye; 37% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 159143 bases at least Q40  
Consensus quality: 159761 bases at least Q30  
Consensus quality: 160127 bases at least Q20  
Insert size: 166000; agarose-fp  
Insert size: 161478; sum-of-ctrls  
Quality coverage: 6.03 in Q20 bases; agarose-fp  
Quality coverage: 6.23 in Q20 bases; sum-of-ctrls

NOTE: This is a working draft sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

*	1	1065:	contig of 1065 bp in length
*	1066	1165:	gap of unknown length
*	1166	32754:	contig of 31589 bp in length
*	32755	32855:	gap of unknown length
*	32855	70034:	contig of 37180 bp in length
*	70035	70134:	gap of unknown length
*	70135	166680:	contig of 90546 bp in length
*	16681	166780:	gap of unknown length
*	166781	161685:	contig of 905 bp in length
*	16186	161785:	gap of unknown length
*	161786	165380:	contig of 795 bp in length.

UNES	Location/Qualifiers
source	1..162580
	/organism="Homo sapiens"
	/dd_xref="taxon:3606"
	/chromosome="YXK"
	/clone="RP11-195B17"
misc_feature	1..1065
	/note="assembly_name:Contig411165..32754"
misc_feature	32835..700034
	/note="assembly_name:Contig4370135..160680"
misc_feature	160781..161665
	/note="assembly_name:Contig41161786..162360"
misc_feature	

	a	c	g	t	others
BASE COUNT	41070	39645	40572	39864	1429
ORIGIN					

Query Match	19.0%;	Score 130.6;	DB 2;	Length 162580;
Best Local Similarity	90.8%;	Pred. No. 4.6e-29;		
Matches 139; Conservative	0;	Mismatches 14;	Indels 0;	Gaps 0

QY	70	ggcgtgaaagggcctctggcgtaccggagataatgcaccatcttcaaggagcgtgcga	129
Db	75995	GGAGTCTCATGTTCTTCTTCTCTTGCGAGATGSCACCAATCTCCAGGAGACGTGCAN	759398
QY	130	aatttttcaaaagatggcctttatgtgtaaagcaacagaaagcctaatgctgcatacgccgt	189
Db	75935	AATTGTGCANAAATGCGCTTTTATTTGTATTAACAGACACAGAGCTAATGCTGCATGCCGT	75876
QY	190	tgcctgcctgaatcagaagggacacatctctgggg	222
Db	75875	TGCTTGCTGATACGAAGGCGCACCTCTTTGGGG	75843

RESULT 10  
AC013413/C

LOCUS	AC013413	192554 bp	DNA	HTG	20-OCT-2000
DEFINITION	Homo sapiens chromosome 2 clone RP11-538J11, WORKING DRAFT				
ACCESSION	AC013413				
VERSION	AC013413.4				
KEYWORDS	GI:8568908				
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 192554)				
JOURNAL	Waterston,R.H.				
REFERENCE	The sequence of Homo sapiens clone				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 192554)				
JOURNAL	Waterston,R.H.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (09-NOV-1999) Genome Sequencing Center, Washington				
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis				
JOURNAL	MO 63108, USA				
COMMENT	On Jun 16, 2000 this sequence version replaced gi:6850521.				

Center: Washington University Genome Sequencing Center	Genome Center
Center code: WUGSC	
Web site: <a href="http://genome.wustl.edu/gsc/index.shtml">http://genome.wustl.edu/gsc/index.shtml</a>	
Project Information	
Center project name: H.NH0538r11	
Summary Statistics	
Sequencing vector: M13: 94%	
Sequencing vector: plasmid: 6%	
Chemistry: dye-primer ET: 94% of reads	
Chemistry: dye-terminator Big Dye: 6% of reads	
Assembly program: Phrap: version 0.990319	
Consensus quality: 175996 bases at least 40	
Consensus quality: 182523 bases at least 30	
Consensus quality: 186265 bases at least 20	
Insert size: 194000; agarose-fp	
Insert size: 191054; sum-of-contigs	
Quality coverage: 3.82 in Q20 bases; agarose-fp	
Quality coverage: 3.82 in Q20 bases; sum-of-contigs	
* NOTE: This is a 'working draft' sequence. It currently	
* consists of 16 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
1	2659: contig of 2659 bp in length
2660	2759: gap of unknown length
2760	5211: contig of 2452 bp in length
5212	5111: gap of unknown length
5312	9206: contig of 3895 bp in length
9207	9306: gap of unknown length
9307	13669: contig of 3763 bp in length
13070	13169: gap of unknown length
13170	17125: contig of 3956 bp in length
17126	17225: gap of unknown length
17226	21986: contig of 4761 bp in length
21987	22086: gap of unknown length
22087	30979: contig of 8893 bp in length
30980	31079: gap of unknown length
31080	40501: contig of 9422 bp in length
40502	40603: gap of unknown length
40602	50933: contig of 10332 bp in length
50934	51033: gap of unknown length
51034	60151: contig of 9118 bp in length
60152	73251: gap of unknown length
73252	73573: contig of 13222 bp in length
73574	73673: gap of unknown length
73674	87260: contig of 13687 bp in length



\* 87261 87360: gap of unknown length  
\* 87361 113485: contig of 26125 bp in length  
\* 113486 113485: gap of unknown length  
\* 113486 138435: contig of 24850 bp in length  
\* 138436 138535: gap of unknown length  
\* 138536 161887: contig of 23352 bp in length  
\* 161888 161987: gap of unknown length  
\* 161988 192554: contig of 30567 bp in length.  
Location/Qualifiers  
1. 192554  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/clone="RP11-538J11"

## FEATURES

source

misc\_feature  
/note="assembly\_name:Contig8"  
2760..5211  
misc\_feature  
/note="assembly\_name:Contig9"  
5312..9206  
misc\_feature  
/note="assembly\_name:Contig10"  
9307..13069  
misc\_feature  
/note="assembly\_name:Contig11"  
13170..17125  
misc\_feature  
/note="assembly\_name:Contig12"  
17226..21986  
misc\_feature  
/note="assembly\_name:Contig13"  
clone\_end:SP6  
vector\_side:right  
22087..30979  
misc\_feature  
/note="assembly\_name:Contig14"  
31080..40501  
misc\_feature  
/note="assembly\_name:Contig15"  
40602..50933  
misc\_feature  
/note="assembly\_name:Contig16"  
51034..60151  
misc\_feature  
/note="assembly\_name:Contig17"  
60252..73473  
misc\_feature  
/note="assembly\_name:Contig18"  
73574..87260  
misc\_feature  
/note="assembly\_name:Contig19"  
87361..113485  
misc\_feature  
/note="assembly\_name:Contig20"  
113586..138435  
misc\_feature  
/note="assembly\_name:Contig21"  
138536..161887  
misc\_feature  
/note="assembly\_name:Contig22"  
161988..192554  
misc\_feature  
/note="assembly\_name:Contig23"  
BASE COUNT 48803 a 46908 c 46387 g 48835 t 1621 others  
ORIGIN

Query Match 19.0% Score 130.6; DB 2; Length 192554;  
Best Local Similarity 90.8%; Pred. No. 4.6e-29;  
Matches 139; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 70 ggagtggaagagggctcgggagctacccgagatgaccccaatgctcagaggaagcgtgcaaa 129  
Db 137112 GGAAGTCTTAAGTTCCTTTCTCTCTGACATATGACACCAATGTCAGGAGCGTGCA 137053  
QY 130 aattgtcaaaagtgagccttattgttaaaagcacagagagcctaagctgacatgcccgt 189  
Db 137052 AATTGTCAAAAGTGAGGCTTTATTGTAAACGACAGAGAGCTAATGTGCAATGCCGT 136993  
QY 190 tgcgtcctgaatcagaagggacacatctctggg 222  
Db 136992 TGCTGCTGATATCAGAAGGCGACCATCTTG6GG 136960

RESULT 11  
AF116241/c 15012 bp DNA PRI 08-JUN-1999  
LOCUS Homo sapiens sodium-dependent multivitamin transporter (SMVT) gene,  
DEFINITION

partial cds.  
ACCESSION AF116241  
VERSION AF116241.1 GI:5006438  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 15012)  
AUTHORS Wang, H., Huang, W., Fei, Y. J., Xia, H., Yang, Feng, T. L., Leibach, F. H., Devore, L. D., Ganapathy, V. and Prasad, P. D.  
TITLE Human placental Na+-dependent multivitamin transporter. Cloning, functional expression, gene structure, and chromosomal localization  
J. Biol. Chem. 274 (21), 14875-14883 (1999)  
JOURNAL MEDLINE  
99262640  
REFERENCE 2 (bases 1 to 15012)  
AUTHORS Prasad, P. D., Wang, H., Leibach, F. H. and Ganapathy, V.  
TITLE Direct Submision  
JOURNAL Submitted (22-DEC-1998) Obstetrics & Gynecology, Medical College of Georgia, 1120 15th Street, Augusta, GA 30912, USA  
FEATURES  
source  
Location/Qualifiers  
1. 15012  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2p23"  
join(<715..898,1396..1462,5141..5672,5989..6054,6397..6448,6838..6905,7427..7581,8000..8140,8342..8471,9067..9155,9589..9701,10054..10121,10866..10952,11093..11274,11487..11590,11827..11942,12361..12504)  
/gene="SMVT"  
/product="sodium-dependent multivitamin transporter"  
<715..>13349  
/gene="SMVT"  
join(5280..5672,5989..6054,6397..6448,6838..6905,7427..7581,8000..8140,8342..8471,9067..9155,9589..9701,10054..10121,10866..10952,11093..11274,11487..11590,11827..11942,12361..12504)  
/gene="SMVT"  
/codon\_start=1  
/product="sodium-dependent multivitamin transporter"  
/protein\_id="A037502.1"  
/db\_xref="GI:5006439"  
/translation="MSYGVSTASPLSPSTGNSVGMSTFSIMDVVEVLLVLSLATGLYHACRGWRHTVGEELIMADRKGCLPVALSLATFGCAVALILGVPSEIFFGQYVFLGCCYFGLILPAHFEIPEYFRIHLTSAEYELERKNTVAVCGTFEIFQMYMGVLYASALNATVGTGDLMSVLALGICTVYTAAGCKKAVIWPVDFOTVILGOLAVIIVGSAKVGGLSRKAVAVASOHORISGFELDDPEYRHTFMTLAFGCPMUSLYGVAQVORLYSRTERKAVALSCYAVPEFQVSLVCGCLIGLVFAYFOEYPMSTIQAKAADQFVLYFVMDLKGJPLGJLGLFIACLFSGSLSTISSAFNSLATVYMEDLIRPMPERSEARAIMLSRGLAFGYGLCLGMAVYSSOMGVLQAVLSIFGMVSGPLGLGFLCMFEPCCANPGAVVGLIAGLVMAFWIGIGSIVTSMGFSMPSPSSGSSFLPTNLVAVATVLLMPLTFESKPTGIGRFRYSLSYLMYSANHSNTVYVGLVSLTGMRGSLNLPATVYPLPKLISLPLSLSCQKRLHCRSYGQDHLIDGLFPEKPRNVGLSDSKEMALDGRAYGSSSTCIIDETS"

BASE COUNT 3179 a 3865 c 4090 g 3878 t  
ORIGIN

Query Match 15.1% Score 103.4; DB 9; Length 15012;  
Best Local Similarity 84.7%; Pred. No. 1e-20;  
Matches 116; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 atggagcctcagagccgggtagcttcagaccctggctgacctggcgccgcctgtgc 60  
Db 549 ATGGAGCCTCAGAGCCGCCGGGTAGCTTTACGACCCCTGGGTGCTGCCGCTGCTC 490  
QY 61 ctgcgtctggcgttgaaaggctctggcgctacccgagatgacccaatgtgcagg 120  
Db 489 CTGCGTCTGGCGTGGAAAGGCTCTGGGCGTACCCGAGGTACAGAACAGTTGAGGT 430

QY 121 agcgtgcaaatgtgtc 137  
| | | | |







```

-----Genome Center
Center: Albert Einstein College of Medicine
Center Code: AECOM
Web site:
http://sequence.aecom.yu.edu/cgi-
bin/ws.exe/mousedB/mousedB/mousedB/sequenceable.hts
Contact: htgs@sequence.aecom.yu.edu
-----Summary Statistics
Center project name: ADD
Sequencing vector: pUC18: L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 202952 at least Q20
*Consensus quality: 202261 at least Q30
*Consensus quality: 200717 at least Q40
*Estimated insert size: agarose-FP - N/A
*Estimated insert size: 208008 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 9.5 x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
30153: contig of 30153 bp in length
30154 30173: gap of unknown length
30174 76426: contig of 46253 bp in length
76427 76446: gap of unknown length
76447 101112: contig of 24666 bp in length
101113 101132: gap of unknown length
101133 123953: contig of 22821 bp in length
123954 123973: gap of unknown length
123974 143713: contig of 19740 bp in length
143714 143733: gap of unknown length
143734 157705: contig of 13972 bp in length
157706 157725: gap of unknown length
157726 167752: contig of 10027 bp in length
167753 167772: gap of unknown length
167773 175672: contig of 7900 bp in length
175673 175692: gap of unknown length
175693 183872: contig of 8180 bp in length
183873 183892: gap of unknown length
183893 189355: contig of 5463 bp in length
189356 189375: gap of unknown length
189376 193648: contig of 4273 bp in length
193649 193668: gap of unknown length
193669 198253: contig of 4585 bp in length
198254 198273: gap of unknown length
198274 200507: contig of 2234 bp in length
200508 200527: gap of unknown length
200528 200973: contig of 446 bp in length
200974 200993: gap of unknown length
200994 202479: contig of 1486 bp in length
202480 202499: gap of unknown length
202500 20367: contig of 868 bp in length
20368 20387: gap of unknown length
20388 204737: contig of 1350 bp in length
204738 204757: gap of unknown length
204758 205655: contig of 898 bp in length
205656 205675: gap of unknown length
205676 206883: contig of 1208 bp in length
206884 206903: gap of unknown length
206904 206988: contig of 85 bp in length
206989 207008: gap of unknown length
207009 207842: contig of 834 bp in length
207843 207862: gap of unknown length
207863 208330: contig of 468 bp in length
208331 208350: gap of unknown length
208351 208448: contig of 96 bp in length.
Location/Qualifiers

```

```

source
1. .208448
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-154K6"
/sex="male"
1. .30153
/note="assembly_name:Contig37"
30174. 76426
/note="assembly_name:Contig36"
76447. 101112
/note="assembly_name:Contig35"
vector_end:T7
vector_side:right
101133. 123953
/note="assembly_name:Contig34"
123974. 143713
/note="assembly_name:Contig33"
143734. 157705
/note="assembly_name:Contig32"
157726. 167752
/note="assembly_name:Contig31"
clone_end:Sp6
vector_side:right
167773. 175672
/note="assembly_name:Contig30"
175693. 183872
/note="assembly_name:Contig29"
183893. 189355
/note="assembly_name:Contig28"
189376. 193648
/note="assembly_name:Contig27"
193669. 198253
/note="assembly_name:Contig26"
198274. 200507
/note="assembly_name:Contig25"
200528. 200973
/note="assembly_name:Contig24"
200994. 202479
/note="assembly_name:Contig23"
202500. 20367
/note="assembly_name:Contig22"
20368. 204737
/note="assembly_name:Contig21"
204758. 205655
/note="assembly_name:Contig20"
205676. 206883
/note="assembly_name:Contig19"
206904. 206988
/note="assembly_name:Contig18"
207009. 207842
/note="assembly_name:Contig17"
207863. 208330
/note="assembly_name:Contig16"
208351. 208448
/note="assembly_name:Contig15"
BASE COUNT 55129 a 49799 c 49920 g 53074 t 526 others
ORIGIN
Query Match 13.4% Score 92: DB 2: Length 208448:
Best Local Similarity 90.7% Pred. No. 4.5e-17:
Matches 98: Conservative 0: Mismatches 10: Indels 0: Gaps 0:
QY 479 gggaccagaatgtgtcctgaagaatgtatctgtgtacctgattgltccaggtctttgc 538
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61079 GGCTCCAGAAAGTGTCTGACAGCGATCTTGATGATGACGCGCTGCTTTGCG 61020
QY 539 agtggtgtgtgtcgtatgtgttccatgatacaagtgatgacccag 586
|||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61019 AGTCGTTGTGCTGATGATGATTCATCGATGATACGAAGTATGAGGACAGG 60972

```







```

Db 189 catgcccgttgcctgctgaatcagagggcaccattctggggtctgattcccaagactgt 248
      |||
Oy 241 tctctgagagaccctgtgccaaattcattcaaggacataaccactgtcatatagaacctg 300
      |||
Db 249 tctctgagagaccctgtgccaaattcattcaaggacataaccactgtcatatagaacctg 308
      |||
Oy 301 caagcaaacccctcaaaagtgtgctggccaacaccttcgtgtgttactaagctccag 360
      |||
Db 309 caagcaaacccctcaaaagtgtgctggccaacaccttcgtgtgttactaagctccag 368
      |||
Oy 361 acctgtactgcacacactgtcaactgtccctggagaaataatccctggaaatactatc 420
      |||
Db 369 acctgtactgcacacactgtcaactgtccctggagaaataatccctggaaatactatc 428
      |||
Oy 421 acctctatatagacaaccaaatctgtcaaggccaagaaccttgcataaactgtgg 480
      |||
Db 429 acctctatatagacaaccaaatctgtcaaggccaagaaccttgcataaactgtgg 488
      |||
Oy 481 gaccagaatagtgtcctgtagaattgtctgtgtactgtatgtccagttccttttgcag 540
      |||
Db 489 gaccagaataatgtgtcctgtagaattgtctgtgtactgtatgtccagttccttttgcag 548
      |||
Oy 541 tgtgttgcctgtagatgttccatgatacaagtgatagtgcagaggtctgtctcaactg 600
      |||
Db 549 tgtgttgcctgtagatgttccatgatacaagtgatagtgcagaggtctgtctcaactg 608
      |||
Oy 601 ctatgtcttctggagattctggagacacacactctatccgtctccatctgtcttggcg 660
      |||
Db 609 ctatgtcttctggagattctggagacacacactctatccgtctccatctgtcttggcg 668
      |||
Oy 661 acccagcgccgaagaaagccaaactca 687
      |||
Db 669 acccagcgccgaagaaagccaaactca 695
      |||

```

## RESULT 3

AAAT85082 standard; DNA: 923 BP.

AAAT85082;

28-JAN-1998 (first entry)

Transforming growth factor alpha HIII polynucleotide.

human transforming growth factor; TGF; TGF-alpha-HIII; angiogenesis;

alopecia; inflammation; ds.

Homo sapiens.

Key Location/Qualifiers

sig\_peptide 5..79

mat\_peptide 80..691

product= TGF-alpha-HIII

/tag= b

/product= TGF-alpha-HIII

W09725349-A1.

17-JUL-1997.

04-JAN-1996; 96WO-US00149.

04-JAN-1996; 96WO-US00149.

(HUMA-) HUMAN GENOME SCI INC.

Well Y;

```

XX WPI: 1997-372817/34.
DR P-PSDB; AAW27087.
XX New human transforming growth factor-alpha homologue - used for
XX developing products for treating e.g. neurological disorders, kidney
XX and liver disorders, tumours, wounds, hair loss or skin disorders
XX Claim 7; Pages 45-46; 63pp; English.
XX This DNA sequence encodes a protein that has been putatively identified
XX as a human transforming growth factor (TGF) alpha analogue,
XX TGF-alpha-HIII. This protein can stimulate angiogenesis, embryogenesis,
XX cell differentiation and function. It can be used for therapeutic
XX purposes for restoration or enhancement of neurological functions
XX diminished as a result of trauma or other damaging pathologies such as
XX AIDS dementia and senile dementia, to treat ocular disorders, e.g.
XX corneal inflammation, to destroy target cells, to treat tumours, kidney
XX or liver disorders or to treat wounds, burns or ulcers. The polypeptide
XX can also be used in the modulation of angiogenesis, bone resorption,
XX immune response, and synaptic and neuronal effector functions, or the
XX arachidonic acid cascade. It can also be used in applications related to
XX terminal differentiation e.g. in hyperproliferative disorders such as
XX inflammation or psoriasis and for alopecia, or other skin conditions
XX which affect hair follicular development. Antagonists to TGF-alpha-HIII
XX can be used for treating tumours or skin disorders such as psoriasis. The
XX products can also be used for diagnosis and detection of the above
XX disorders.
SQ Sequence 923 BP; 235 A; 240 C; 210 G; 238 T; 0 other:

```

Query Match 100.0%; Score 687; DB 18; Length 923;

Best Local Similarity 100.0%; Pred. No. 1,1e-210; Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 atggcgctcctacgcccgggtgattcttaccgacctgtgtccctgggtcgcgcctctgc 60
      |||
Db 5 atggcgctcctacgcccgggtgattcttaccgacctgtgtccctgggtcgcgcctctgc 64
      |||
Oy 61 ctgcgtctgggtggaaggggtctggcgctaccggagatgtgacccaatgtccagg 120
      |||
Db 65 ctgcgtctgggtggaaggggtctggcgctaccggagatgtgacccaatgtccagg 124
      |||
Oy 121 agcgtgcaaatgtgcaaaagtgtgcttattgttaaacgacacgagagtaatgtc 180
      |||
Db 125 agcgtgcaaatgtgcaaaagtgtgcttattgttaaacgacacgagagtaatgtc 184
      |||
Oy 181 catgcccgtgtcctgtgaatcagaagggcacacacttggggctgtgattccagaactgt 240
      |||
Db 185 catgcccgtgtcctgtgaatcagaagggcacacacttggggctgtgattccagaactgt 244
      |||
Oy 241 tctctgagagaccctgtccaaattcattcaggaacataaccactgtcatatagaacctg 300
      |||
Db 245 tctctgagagaccctgtccaaattcattcaggaacataaccactgtcatatagaacctg 304
      |||
Oy 301 caagcaaacccctcaaaagtgtgctggccaacaccttcgtgtgttactaagctccag 360
      |||
Db 305 caagcaaacccctcaaaagtgtgctggccaacaccttcgtgtgttactaagctccag 364
      |||
Oy 361 acctgtactgcacacactgtcaactgtccctggagaaataatccctggaaatactatc 420
      |||
Db 365 acctgtactgcacacactgtcaactgtccctggagaaataatccctggaaatactatc 424
      |||
Oy 421 acctctatatagacaaccaaatctgtcaaggccaagaaccttgcataaactgtgg 480
      |||
Db 425 acctctatatagacaaccaaatctgtcaaggccaagaaccttgcataaactgtgg 484
      |||
Oy 481 gaccagaataatgtgtcctgtagaattgtgtatccctgtagtccaggtcttggcg 540
      |||
Db 485 gaccagaataatgtgtcctgtagaattgtgtatccctgtagtccaggtcttggcg 544
      |||
Oy 541 tgtgttgcctgtagtcttccatgatacaagtgatgtgcagaggtctgtctcaactg 600
      |||

```

```

Db      545 tgtgttgcgcgatggttccatgatacaagttatgagccgaggtcgttccactg 604
Qy      601 ctatgttcttcggattctggagaccactctatccgtctccattcgtttggcg 660
Db      605 ctatgttcttcggattctggagaccactctatccgtctccattcgtttggcg 664
Qy      661 acccagcgccgaaagccaagacttca 687
Db      665 acccagcgccgaaagccaagacttca 691

RESULT 4
AAS08543
ID      AAS08543 standard: cDNA: 923 BP.
XX
XX      AAS08543:
XX
XX      24-OCT-2001 (first entry)
XX
XX      DNA encoding transforming growth factor (TGF) alpha HIII.
XX
XX      Human; TGF alpha HIII; transforming growth factor alpha HIII; cancer;
XX      diagnostic; therapeutic; immune disorder; multiple sclerosis;
XX      systemic lupus erythematosus; human immuno-deficiency virus; HIV;
XX      hyperproliferative disorder; Gaucher's disease; cardiovascular disease;
XX      Schmitz syndrome; Chaga's cardiomyopathy; coronary arteriosclerosis;
XX      angiogenic disorder; corneal graft; neovascularisation; wound healing;
XX      diabetic retinopathy; neurological disorder; Huntington's chorea;
XX      Alzheimer's disease; Parkinson's disease; ss.
XX
XX      Homo sapiens.
XX
XX      Key      Location/Qualifiers
XX      CDS      5..694
XX              /*tag= a
XX              /product= "TGF alpha HIII"
XX              /note= "Transforming growth factor (TGF) alpha HIII"
XX              sig_peptide
XX              5..79
XX              /*tag= b
XX              mat_peptide
XX              80..691
XX              /*tag= c
XX              /note= "Mature TGF alpha HIII"
XX
XX      MO200140251-A1.
XX
XX      07-JUN-2001.
XX
XX      01-DEC-2000; 2000WO-US32745.
XX
XX      02-DEC-1999; 99US-0168387.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      Wei Y;
XX
XX      WPI: 2001-441480/47.
XX      P-PSDB: AAU04295.
XX
XX      Nucleic acid encoding human transforming growth factor alpha III
XX      (TGFa), useful for preventing, diagnosing and/or treating e.g. Cancer
XX      and Parkinson's disease -
XX
XX      Claim 1; Fig 1; 302pp; English.
XX
XX      The sequence represents the coding sequence of human transforming
XX      growth factor (TGF) alpha HIII. TGF alpha HIII nucleic acid and protein
XX      may be used in the prevention, diagnosis and treatment of diseases
XX      associated with inappropriate polypeptide expression, for example immune
XX      disorders (e.g. multiple sclerosis, systemic lupus erythematosus and
XX      human immuno-deficiency virus (HIV) infections), hyperproliferative
XX      disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
XX      (e.g. Schmitz syndrome, Chaga's cardiomyopathy and coronary

```

```

CC      arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC      neovascularisation and diabetic retinopathy), neurological disorders
CC      (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC      infectious diseases and/or for promoting wound healing, regeneration
CC      and/or chemotaxis (full details given in specification). Additionally,
CC      the nucleic acid may be used to produce the secreted polypeptides, by
CC      inserting the nucleic acids into a host cell and culturing the cell to
CC      express the protein. It may also be used as a DNA probe in diagnostic
CC      assays to detect and quantitate the presence of similar nucleic acid
CC      sequences in samples, and therefore which patients may be in need of
CC      restorative therapy. The polypeptides may also be used as antigens in the
CC      production of antibodies against TGF alpha HIII and in assays to identify
CC      modulators of TGF alpha HIII. The anti-TGF alpha HIII antibodies may also
CC      be used as diagnostic agents for detecting the presence of TGF alpha HIII
CC      in samples (e.g. by enzyme linked immunosorbent assay (ELISA)).
XX
XX      Sequence 923 BP; 235 A; 240 C; 210 G; 238 T; 0 other;
XX
XX      Query Match      100.0%; Score 687; DB 22; Length 923;
XX      Best Local Similarity 100.0%; Pred. No. 1,1e-210;
XX      Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 atggcgctcagcgccggtagtcttaccagccctgtgcccgtggcctgcgcctc 60
Db      5 atggcgctcagcgccggtagtcttaccagccctgtgcccgtggcctgcgcctc 64
Qy      61 ctgcgtctggcgctggaaaggctctgcgcgtaccggagatatgcacccaatgtccagg 120
Db      65 ctgcgtctggcgctggaaaggctctgcgcgtaccggagatatgcacccaatgtccagg 124
Qy      121 agcgctgcaaatgttgcacaaagtgagcctttattgaaacagacagaaagtaatgctg 180
Db      125 agcgctgcaaatgttgcacaaagtgagcctttattgaaacagacagaaagtaatgctg 184
Qy      181 catgccgttgcctgccttgatcagaaggacacacttgggctggtgatctccagaactc 240
Db      185 catgccgttgcctgccttgatcagaaggacacacttgggctggtgatctccagaactc 244
Qy      241 tctctgagagaccctgtgccaaactttcatcagacacatacactgtcatcactg 300
Db      245 tctctgagagaccctgtgccaaactttcatcagacacatacactgtcatcactg 304
Qy      301 caagcaaacccctcaaaagtgtacttgagcaaaccttccgtggttactaactccag 360
Db      305 caagcaaacccctcaaaagtgtacttgagcaaaccttccgtggttactaactccag 364
Qy      361 actctgatactgcaacaactgtcaactgtccttgaggaataatgacctgatactatc 420
Db      365 actctgatactgcaacaactgtcaactgtccttgaggaataatgacctgatactatc 424
Qy      421 acctctatatagacaacaactgttcaaggcgaaagaaccttgcataaactgg 480
Db      425 acctctatatagacaacaactgttcaaggcgaaagaaccttgcataaactgg 484
Qy      481 gaccagaatgtgtcttgaagtgaatcttgaacctgagtcgagtcgagtcctttgag 540
Db      485 gaccagaatgtgtcttgaagtgaatcttgaacctgagtcgagtcgagtcctttgag 544
Qy      541 tgtgttgcctgtagtgttccatgatacaagttatgagccgaggtcgttccactg 600
Db      545 tgtgttgcctgtagtgttccatgatacaagttatgagccgaggtcgttccactg 604
Qy      601 ctatgttcttcggattctggagaccactctatccgtctccattcgtttggcg 660
Db      605 ctatgttcttcggattctggagaccactctatccgtctccattcgtttggcg 664
Qy      661 acccagcgccgaaagccaagacttca 687
Db      665 acccagcgccgaaagccaagacttca 691

RESULT 5

```



AA58584  
 ID AAC58584 standard; CDNA: 932 BP.  
 XX  
 AC AAC58584;  
 XX  
 DT 29-JAN-2001 (first entry)  
 XX  
 DE Human PRO240 protein UNQ214 encoding cDNA SEQ ID NO:25.  
 XX  
 XX Human; immune related disease; diagnosis; antiinflammatory; cardiant;  
 KW dermatological; antiaarthritis; antirheumatic; immunosuppressive;  
 KW haemostatic; antihypoid; antidiabetic; nootropic; neuroprotective;  
 KW antiaarthritis; hepatotropic; virucide; antiparasitic; antiallergic;  
 KW osteoarthritis; systemic lupus erythematosus; rheumatoid arthritis;  
 KW idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;  
 KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;  
 KW autoimmune thrombocytopaenia; immune-mediated renal disease;  
 KW demyelinating disease; hepatobiliary disease; Whipple's disease;  
 KW inflammatory bowel disease; gluten-sensitive enteropathy;  
 KW autoimmune disease; immune-mediated skin disease; allergic disease;  
 KW immunological disease; transplantation associated disease;  
 KW graft rejection; graft-versus-host-disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX MO200053758-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PE 02-MAR-2000; 2000MO-US05841.  
 XX  
 XX 08-MAR-1999; 99WO-US05028.  
 PR 10-MAR-1999; 99US-0123618.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 23-MAR-1999; 99US-0125775.  
 PR 12-APR-1999; 99US-0128849.  
 PR 20-APR-1999; 99WO-US08615.  
 PR 28-APR-1999; 99US-0131445.  
 PR 04-MAY-1999; 99US-0132371.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144738.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 28-JUL-1999; 99US-0146222.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21547.  
 PR 05-OCT-1999; 99WO-US23089.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 29-NOV-1999; 99WO-US28214.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 30-NOV-1999; 99WO-US28409.  
 PR 01-DEC-1999; 99WO-US28301.  
 PR 02-DEC-1999; 99WO-US28634.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28564.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 20-DEC-1999; 99WO-US31274.  
 PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000MO-US00219.  
 PR 06-JAN-2000; 2000MO-US00277.  
 PR 11-FEB-2000; 2000MO-US00376.  
 PR 18-FEB-2000; 2000MO-US04341.  
 PR 18-FEB-2000; 2000MO-US04342.  
 PR 22-FEB-2000; 2000MO-US04414.  
 XX  
 PA (GENE) GENENTECH INC.

XX  
 PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W;  
 PI Kabakoff RC, Lu Y, Pan J, Pennica D, Shelton DL, Smith V;  
 PI Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M;  
 XX  
 XX WPI: 2000-572271/53.  
 DR P-PSDB; AAB33419.  
 DR  
 DR  
 XX  
 PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of  
 PT immune related disorders, e.g. systemic lupus erythematosus, rheumatoid  
 PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -  
 XX  
 PS Claim 23; Fig 11; 309pp; English.  
 CC The present invention describes sixty four human PRO proteins which can  
 CC be used in the treatment of immune related diseases. The human PRO  
 CC proteins, anti-PRO antibodies, agonists and antagonists are useful for  
 CC treating and diagnosing immune related disorders. The disorders are  
 CC selected from systemic lupus erythematosus, rheumatoid arthritis,  
 CC osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,  
 CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's  
 CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic  
 CC anaemia, autoimmune thrombocytopaenia, thyroiditis, diabetes mellitus,  
 CC immune-mediated renal disease, demyelinating diseases of the central  
 CC and peripheral nervous systems, hepatobiliary diseases, inflammatory  
 CC bowel disease, gluten-sensitive enteropathy and Whipple's disease,  
 CC autoimmune or immune-mediated skin diseases, allergic diseases,  
 CC immunological diseases of the lung, and transplantation associated  
 CC diseases including graft rejection and graft-versus-host-disease.  
 CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used  
 CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and protein  
 CC AAB33414 to AAB33477 represent human PRO polynucleotide and protein  
 CC sequences given in the exemplification of the present invention.  
 XX  
 SQ Sequence 932 BP; 239 A; 241 C; 214 G; 237 T; 1 other;  
 Query Match 100.0%; Score 687; DB 21; Length 932;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-210;  
 Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 atggcgcccaagcggcggtgcttcaagccctgtgcccggcgctgcgc 60  
 DB 12 atggcgcccaagcggcggtgcttcaagccctgtgcccggcgctgc 71  
 QY 61 ctgcgtctggcgctggaaggcgctgctgcttcaagccgagatgcaacccaaatgccaagg 120  
 DB 72 ctgcgtctggcgctggaaggcgctgctgcttcaagccgagatgcaacccaaatgccaagg 131  
 QY 121 agcgctgcaaatgttgcacaagtgtgcttattgtgaaaaagcagaagagctaatgctg 180  
 DB 132 agcgctgcaaatgttgcacaagtgtgcttattgtgaaaaagcagaagagctaatgctg 191  
 QY 181 catgcccgttgcctgctgaatcagaaggcaccattgtggcgctggtatccagaactgt 240  
 DB 192 catgcccgttgcctgctgaatcagaaggcaccattgtggcgctggtatccagaactgt 251  
 QY 241 tctctggaagaccctgtgccaacttcaacagcacaataccactgcatcatagaactgt 300  
 DB 252 tctctggaagaccctgtgccaacttcaacagcacaataccactgcatcatagaactgt 311  
 QY 301 caagcaaacccctcaagtgactgtgccaacacttccgtgtgcttactagctgcag 360  
 DB 312 caagcaaacccctcaagtgactgtgccaacacttccgtgtgcttactagctgcag 371  
 QY 361 actctgatactgcacaactgtcaactgtctggaaggaataatgctgtaatactatc 420  
 DB 372 actctgatactgcacaactgtcaactgtctggaaggaataatgctgtaatactatc 431  
 QY 421 acctctatatagacaacaaatctgtcaagggaagaacacttgcataaactgga 480  
 DB 432 acctctatatagacaacaaatctgtcaagggaagaacacttgcataaactgga 491

QY 481 gaccagaatgtgtctcagaaatgatctgttaccatgagtcgaagctcttgcag 540  
 |||||||  
 Db 492 gaccagaatgtgtctcagaaatgatctgttaccatgagtcgaagctcttgcag 551  
 QY 541 tgtgttctgtcgtatggttccatgatacaagtatgagccaggctcgttccactg 600  
 |||||||  
 Db 552 tgtgttctgtcgtatggttccatgatacaagtatgagccaggctcgttccactg 611  
 QY 601 ctatgttctgtcgtatggttccatgatacaagtatgagccaggctcgttccactg 660  
 |||||||  
 Db 612 ctatgttctgtcgtatggttccatgatacaagtatgagccaggctcgttccactg 671  
 QY 661 acccagcgcgaaagccaagacttca 687  
 |||||||  
 Db 672 acccagcgcgaaagccaagacttca 698  
  
 RESULT 6  
 AAA30036 standard; cDNA; 932 BP.  
 XX  
 AC AAA30036;  
 XX  
 DT 09-AUG-2000 (first entry)  
 XX  
 DE Human PRO240 nucleotide sequence.  
 XX  
 KW Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;  
 KW PRO17; tumour growth inhibitor; cancer; diagnosis; treatment; human;  
 KW cell growth; proliferation; serrate precursor; C-serrate-1; ADPPT;  
 KW antibody dependent enzyme mediated prodruq therapy; chromosome 2; ss.  
 OS Homo sapiens.  
 PN WO200015666-A2.  
 XX  
 PD 23-MAR-2000.  
 XX  
 PF 08-SEP-1999; 99WO-US20594.  
 XX  
 PR 10-SEP-1998; 98US-0099803.  
 PR 10-SEP-1998; 98WO-US18824.  
 XX  
 PA (GENTH ) GENENTECH INC.  
 PI Goddard A, Gurney AL, Hillan KJ, Roy MA, Wood WL, Botstein D;  
 XX  
 DR WPI: 2000-271386/23.  
 DR P-PSDB: AAY88569.  
 XX  
 PT New isolated antibodies which bind to specific polypeptides used for  
 XX diagnosis and treatment of neoplastic cell growth and proliferation -  
 PS Example 4; Fig 7; 200pp; English.  
 XX  
 CC This sequence represents a human PRO240 nucleotide sequence. PRO240  
 CC shares sequence homology with the D. melanogaster serrate precursor  
 CC protein and the Gallus gallus C-serrate-1 protein. The PRO240 gene is  
 CC located on chromosome 2. The invention relates to isolated antibodies  
 CC which bind to a polypeptide. The "PRO" polypeptides are encoded by genes  
 CC which are over expressed in the genome of tumour cells. Vectors and host  
 CC cells comprising the nucleic acid encoding the antibodies are used in the  
 CC production of the antibodies. The antibodies and nucleic acids encoding  
 CC them are used for diagnosing a tumour in a mammal. The antibodies are  
 CC used for inhibiting the growth of tumour cells and identifying compounds  
 CC that inhibit a biological or immunological activity of and/or expression  
 CC of a PRO187, PRO533, PRO214, PRO240, PRO211, PRO230, PRO261, PRO246 or  
 CC PRO17 polypeptide. The antibody can be used in antibody dependent enzyme  
 CC mediated prodruq therapy (ADPPT) by conjugating the antibody to a  
 CC prodruq-activating enzyme which converts a prodruq to an anti-cancer  
 CC drug. The antibodies can be fluorescently labeled and monitored by light  
 CC microscopy, flow cytometry or fluorimetry for diagnosis and prognosis of  
 CC tumours.

XX  
 SQ Sequence 932 BP: 239 A; 241 C; 214 G; 237 T; 1 other;  
  
 Query Match 100.0%; Score 687; DB 21; Length 932;  
 Best Local Similarity 100.0%; Pred. No. 1, 1e-210;  
 Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 atgagcgtccacggccgggtaagtcttaagacacctgtgtccctgggtcggccctgc 60  
 |||||||  
 Db 12 atggcgtccacggccgggtaagtcttaagacacctgtgtccctgggtcggccctgc 71  
 QY 61 ctgcctctggcgctggaaaggcctctgctgcgtaccacagatagcaccatgtccaagg 120  
 |||||||  
 Db 72 ctgcctctggcgctggaaaggcctctgctgcgtaccacagatagcaccatgtccaagg 131  
 QY 121 agcgtlgaanaattgtcaaaagtggccttattgtlaaaagacagagagactaagctg 180  
 |||||||  
 Db 132 agcgtlgaanaattgtcaaaagtggccttattgtlaaaagacagagagactaagctg 191  
 QY 181 catgcccgttgcctcgtatcagaaggacacatcttgggcttgatctccagaactgt 240  
 |||||||  
 Db 192 catgcccgttgcctcgtatcagaaggacacatcttgggcttgatctccagaactgt 251  
 QY 241 tctctggagagacctgtgtccaacttcatcagacacatcctgtcatcatagacctg 300  
 |||||||  
 Db 252 tctctggagagacctgtgtccaacttcatcagacacatcctgtcatcatagacctg 311  
 QY 301 caagcaaacccctcaaaagtgacttggccaacaccttccgtgcttactcagctccag 360  
 |||||||  
 Db 312 caagcaaacccctcaaaagtgacttggccaacaccttccgtgcttactcagctccag 371  
 QY 361 actctgtactcgcacacacatgtcaactgtctcggagggaattatcctcgtgaattatc 420  
 |||||||  
 Db 372 actctgtactcgcacacacatgtcaactgtctcggagggaattatcctcgtgaattatc 431  
 QY 421 acctctatatagacaaccaatctgtcaaggacaaagacctgtgacaatacactggg 480  
 |||||||  
 Db 432 acctctatatagacaaccaatctgtcaaggacaaagacctgtgacaatacactggg 491  
 QY 481 gaccagaatgtgtctcagaaatgatctgtgtactgtatgtgtccaggtcctttgcag 540  
 |||||||  
 Db 492 gaccagaatgtgtctcagaaatgatctgtgtactgtatgtgtccaggtcctttgcag 551  
 QY 541 tgtgttctgtcgtatggttccatgatacaagtatgagccaggctcgttccactg 600  
 |||||||  
 Db 552 tgtgttctgtcgtatggttccatgatacaagtatgagccaggctcgttccactg 611  
 QY 601 ctatgttctgtcgtatggttccatgatacaagtatgagccaggctcgttccactg 660  
 |||||||  
 Db 612 ctatgttctgtcgtatggttccatgatacaagtatgagccaggctcgttccactg 671  
 QY 661 acccagcgcgaaagccaagacttca 687  
 |||||||  
 Db 672 acccagcgcgaaagccaagacttca 698  
  
 RESULT 7  
 AAF30054 standard; cDNA; 932 BP.  
 XX  
 ID AAF30054  
 XX  
 AC AAF30054;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Human cDNA encoding PRO240.  
 XX  
 KW PRO240; UNO214; human; immune disease; autoimmune disease;  
 KW antihemmatic; antithyroid; antinflammatory; antianaemic;  
 KW immunosuppressive; antidiabetic; neuroprotective;  
 KW hepatocytic; virucide; dermatological; antiproliferative;  
 KW antiallergic; immunostimulant; serrate;  
 KW lung cancer; ss.

XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 12..701  
 FT sig\_peptide /\*tag= a  
 FT 12..101  
 FT mat\_peptide /\*tag= b  
 FT 102..698  
 FT /\*tag= c  
 XX  
 PN WO200105972-A1.  
 PD 25-JAN-2001.  
 PE 15-MAR-2000; 2000MO-US06884.  
 PR 20-JUL-1999; 99US-0144/58.  
 PA (GENTH ) GENENTECH INC.  
 XX  
 PI Ashkenazi AJ, Baker KP, Fong S, Goddard A, Godowski PJ, Gurney AL;  
 PI Hillan KU, Mark MR, Masters SA, Pitti RM, Tumas D, Wetanabe CK;  
 PI Wood WJ;  
 DR WPI: 2001-103149/11.  
 DR P-PSDB; AAB20112.  
 XX  
 PT New PRO polypeptides, nucleic acids and (ant)agonists, useful for  
 PT diagnosing and treating immune-related disorders, such as multiple  
 PT sclerosis, rheumatoid arthritis and diabetes -  
 XX  
 PS Claim 21; Fig 9; 127pp; English.  
 CC  
 CC The present sequence is that of cDNA clone DNR34387-1138 (ATCC 209260)  
 CC encoding novel human immunomodulator protein PRO240 (UNQ214) (see  
 CC AAB20112). The clone was isolated from a foetal liver tissue cDNA  
 CC library. The predicted protein (25 kDa, pI 7.83) shows homology to  
 CC Drosophila serate precursor and chicken C-serate-1. Expression  
 CC was observed in lung cancer, 8 squamous carcinomas and in 6/8  
 CC adenocarcinomas, in situ and infiltrating components. The invention  
 CC provides polynucleotides (see AAF30050-62) encoding novel human PRO  
 CC proteins (see AAB20108-20) including PRO240. Claimed compositions  
 CC comprising these proteins or their agonists are useful for increasing  
 CC infiltration of inflammatory cells into a tissue of a mammal,  
 CC stimulating or enhancing an immune response in a mammal, or  
 CC increasing the proliferation of T-lymphocytes in a mammal in response  
 CC to an antigen. Claimed compositions comprising the PRO polypeptide  
 CC or its antagonist have the opposite effect. A claimed method for  
 CC treating an immune related disorder, such as a T cell disorder,  
 CC involves administering the PRO polypeptide, an agonist antibody or  
 CC an antagonist antibody. The disorder is selected from systemic  
 CC lupus erythematosus, rheumatoid arthritis, osteoarthritis,  
 CC juvenile chronic arthritis, spondyloarthritis, systemic  
 CC sclerosis, idiopathic inflammatory myopathy, Sjogren's syndrome,  
 CC systemic vasculitis, sarcoidosis, autoimmune haemolytic anaemia,  
 CC autoimmune thrombocytopenia, thyroiditis, diabetes mellitus,  
 CC immune-mediated renal disease, demyelinated diseases (such as  
 CC multiple sclerosis), autoimmune chronic active hepatitis, primary  
 CC biliary cirrhosis, granulomatous hepatitis, sclerosing cholangitis,  
 CC inflammatory bowel disease (ulcerative colitis and Crohn's disease),  
 CC gluten-sensitive enteropathy, Whipple's disease, (auto)immune-mediated  
 CC skin diseases (such as bullous skin disease, erythema multiforme and  
 CC psoriasis), allergic diseases (such as asthma, allergic rhinitis,  
 CC atopic dermatitis, food hypersensitivity and urticaria), immunologic  
 CC diseases of the lung and transplantation associated diseases (such  
 CC as graft rejection and graft-versus-host disease) (all claimed).  
 CC Claimed methods of diagnosing these disorders comprise detecting  
 CC the level of expression of the PRO gene. Also claimed are a method  
 CC of identifying a compound capable of inhibiting the expression or  
 CC activity of the PRO polypeptide, vectors, host cells, antibodies  
 CC and a method of stimulating an immune response by administering  
 CC PRO240.

XX  
 SQ Sequence 932 BP; 239 A; 241 C; 214 G; 237 T; 1 other;  
 Query Match 100.0%; Score 687; DB 22; Length 932;  
 Best Local Similarity 100.0%; Pred. No. 1..e-210;  
 Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 atggcgctcagcgccgggtagcttcaagaccggtgccccgggtgccccctgtc 60  
 DB 12 atggcgctcagcgccgggtagcttcaagaccggtgccccgggtgccccctgtc 71  
 QY 61 ctgcctctggcgctggaagggtctggtgctaccccgagatagaccacgaatccagg 120  
 DB 72 ctgcctctggcgctggaagggtctggtgctaccccgagatagaccacgaatccagg 131  
 QY 121 agcgtgcaaaattgtcaaaagtgcctttatgttaaaagacagagagtaagtctg 180  
 DB 132 agcgtgcaaaattgtcaaaagtgcctttatgttaaaagacagagagtaagtctg 191  
 QY 181 catgcccgttgcctgtgatacgaaggcaccatcttggtggctgatacagaaactgt 240  
 DB 192 catgcccgttgcctgtgatacgaaggcaccatcttggtggctgatacagaaactgt 251  
 QY 241 tctctggagagacctgtgtcaaaacttcaagacacatacaccatgtcatatagactg 300  
 DB 252 tctctggagagacctgtgtcaaaacttcaagacacatacaccatgtcatatagactg 311  
 QY 301 caagcaaacccctcaaaagtgtacttggcaaaccttcctgtgcttactaagtcag 360  
 DB 312 caagcaaacccctcaaaagtgtacttggcaaaccttcctgtgcttactaagtcag 371  
 QY 361 acctatctgtccacaacatgtcaactgtctctgtagagaaatgaagcgaatactac 420  
 DB 372 acctatctgtccacaacatgtcaactgtctctgtagagaaatgaagcgaatactac 431  
 QY 421 acctctatatagacaacaaatctgtcaaggcacaagaaccttgcataaactggg 480  
 DB 432 acctctatatagacaacaaatctgtcaaggcacaagaaccttgcataaactggg 491  
 QY 481 gaccagaatgtgtcctgaagaatgtatctgtacactgtatgtccaggtctttgcag 540  
 DB 492 gaccagaatgtgtcctgaagaatgtatctgtacactgtatgtccaggtctttgcag 551  
 QY 541 tgggttggctgatatgtttcatgatacaagtgatgagcgaaggtcgttccagct 600  
 DB 552 tgggttggctgatatgtttcatgatacaagtgatgagcgaaggtcgttccagct 611  
 QY 601 ctatgttcttcggagattctgtaggacacacactatccgttccattctgttggcg 660  
 DB 612 ctatgttcttcggagattctgtaggacacacactatccgttccattctgttggcg 671  
 QY 661 acccagcgccgaagaagcgaagactca 687  
 DB 672 acccagcgccgaagaagcgaagactca 698  
 RESULT 8  
 AAF60356  
 ID AAF60356 standard; cDNA; 932 BP.  
 XX  
 AC AAF60356;  
 XX  
 DT 27-APR-2001 (first entry)  
 XX  
 DE PRO240 coding sequence.  
 XX  
 KW Cytostatic; PRO protein; tumour; cancer; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200105836-A1.  
 XX

PD 25-JAN-2001.  
 XX 20-DEC-1999; 99WO-US30999..  
 XX 20-JUL-1999; 99US-0144758.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 05-OCT-1999; 99WO-US23089.  
 PR 29-NOV-1999; 99WO-US28234.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 02-DEC-1999; 99WO-US28564.  
 XX  
 XX (GETH ) GENENTECH INC.  
 PA  
 PI Bostein D, Goddard A, Gurney AL, Hillian KJ, Roy MA, Wood WI;  
 XX  
 DR WPI; 2001-091968/10.  
 DR P-PSDB; AAB68595.  
 XX  
 PT New antibody that binds to a PRO polypeptide, e.g. PRO187 and PRO533,  
 PS useful for diagnosing and treating cancers -  
 XX  
 PS Claim 50; Fig 7; 196pp; English.  
 XX  
 CC The present invention relates to PRO proteins and coding sequences. The  
 CC present sequence is the coding sequence for one such PRO protein.  
 CC It was found that the PRO genes are amplified in the genome of tumour  
 CC cells. The gene amplification is expected to be associated with the  
 CC overexpression of the gene product and contributes to tumorigenesis.  
 CC Therefore, antagonists of PRO proteins are useful for the treatment of  
 CC benign or malignant tumours, leukaemias, lymphoid malignancies and other  
 CC disorders such as neuronal, glial, astrocytic, hypothalamic, glandular,  
 CC epithelial, inflammatory and immunologic disorders.  
 CC  
 XX Sequence 932 BP; 239 A; 241 C; 214 G; 237 T; 1 other;  
 SQ

Query Match 100.0%; Score 687; DB 22; Length 932;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-210;  
 Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcgccctcaagcgccggtatgtctttagacccctggtccctggtccgctgtc 60  
 |||||  
 DB 12 atggcgccctcaagcgccggtatgtctttagacccctggtccctggtccgctgtc 71  
 QY 61 ctgcgtctgggctgtgaaaggctctggtcgtaccgagatagcaccatgtccagg 120  
 |||||  
 DB 72 ctgcgtctgggctgtgaaaggctctggtcgtaccgagatagcaccatgtccagg 131  
 QY 121 agcggtgcaaatgttgcataaagtgctcttattgttaaagcaagagagctaatgctg 180  
 |||||  
 DB 132 agcggtgcaaatgttgcataaagtgctcttattgttaaagcaagagagctaatgctg 191  
 QY 181 catgcccgttctgctgctaatcagaagggaccatcttggggctgagatctcagaactgt 240  
 |||||  
 DB 192 catgcccgttctgctgctaatcagaagggaccatcttggggctgagatctcagaactgt 251  
 QY 241 tctctgagagaccctgtgtcacaactttcacaagcaatatacactgtatcatatgacctg 300  
 |||||  
 DB 252 tctctgagagaccctgtgtcacaactttcacaagcaatatacactgtatcatatgacctg 311  
 QY 301 caagcaaaacccctcaaaaggtgacttgcacaacaccttcgtgcttactacagctccag 360  
 |||||  
 DB 312 caagcaaaacccctcaaaaggtgacttgcacaacaccttcgtgcttactacagctccag 371  
 QY 361 actcgtatctgcacacacatgtcaactgtctctgagagaaataatgctcgtgaatactatc 420  
 |||||  
 DB 372 actcgtatctgcacacacatgtcaactgtctctgagagaaataatgctcgtgaatactatc 431  
 QY 421 accttctatagacaaccaaactgttcaaggcaaaagcaacttgcataaactgtgg 480  
 |||||

DB 432 acctcttatagacaaccaaatctgtcaaggcgaagaacaccttgcataaactgtgg 491  
 QY 481 gaccagaataatgtctctgaagaatgatcttgtatccctgatgtccaggtctttggcg 540  
 |||||  
 DB 492 gaccagaataatgtctctgaagaatgatcttgtatccctgatgtccaggtctttggcg 551  
 QY 541 tgtgtttgtgtgatgtgttccatagatacaagtgatgtgcgaaggtcgttctcaactg 600  
 |||||  
 DB 552 tgtgtttgtgtgatgtgttccatagatacaagtgatgtgcgaaggtcgttctcaactg 611  
 QY 601 ctatgttcttctggagatcttggagccaccactctatccgttccatcttcttggcg 660  
 |||||  
 DB 612 ctatgttcttctggagatcttggagccaccactctatccgttccatcttcttggcg 671  
 QY 661 acccagcgccgaagaagccaagacttca 687  
 |||||  
 DB 672 acccagcgccgaagaagccaagacttca 698  
 |||||  
 RESULT 9  
 AAC91553  
 ID AAC91553 standard; cDNA; 932 BP.  
 XX  
 AC AAC91553;  
 XX  
 DT 21-MAR-2001 (first entry)  
 XX  
 DE Human PRO240 cDNA.  
 XX  
 KW Human: PRO; cytosolic; neotropic; neuroprotective; respiratory general;  
 KW antiinflammatory; antiangiogenic; immunosuppressive; immunostimulant;  
 KW PRO agonist; cancer; inflammatory disorder; immunological disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200073348-A2.  
 PD 07-DEC-2000.  
 XX  
 PF 30-MAY-2000; 2000WO-US14941.  
 XX  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 22-JUN-1999; 99US-0140650.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 01-DEC-1999; 99WO-US28634.  
 PR 02-DEC-1999; 99WO-US28651.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 20-DEC-1999; 99WO-US30999.  
 PR 06-JAN-2000; 2000WO-US00376.  
 PR 11-FEB-2000; 2000WO-US03565.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 02-MAR-2000; 2000WO-US05841.  
 PR 03-MAR-2000; 2000US-0187202.  
 PR 10-MAR-2000; 2000WO-US06319.  
 PR 15-MAR-2000; 2000WO-US06884.  
 PR 30-MAR-2000; 2000WO-US08439.  
 PR 17-MAY-2000; 2000WO-US13705.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W, Kabakoff RC;  
 PI Shelton DL, Smith V, Watanabe CK, Wood WI;  
 XX  
 DR WPI; 2001-016509/02.  
 DR P-PSDB; AAB50951.  
 XX  
 PT Twenty eight nucleic acids encoding PRO polypeptides which are useful



CC acid molecules encoding the EXCS of the invention.  
 XX Sequence 1265 BP; 301 A; 341 C; 356 G; 267 T; 0 other;  
 SQ

Query Match 100.0%; Score 687; DB 22; Length 1265;  
 Best Local Similarity 100.0%; Pred. No. 1,3e-210;  
 Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggcgctcaagcgccggtgagttcttcagaccctgtgcttggtgctgcgcctgtc 60  
 DB |||||||  
 QY 355 atggcgctcaagcgccggtgagttcttcagaccctgtgcttggtgctgcgcctgtc 414  
 DB |||||||  
 QY 61 ctgcgtctgggagcgcttgagcgcttcacccgagatcaccccaatgctcagg 120  
 DB |||||||  
 QY 415 ctgcgtctgggagcgcttgagcgcttcacccgagatcaccccaatgctcagg 474  
 DB |||||||  
 QY 121 agcgtgcaaaattgttcaaaagtgtgctttatgttaaacgacagagactaatgctg 180  
 DB |||||||  
 QY 475 agcgtgcaaaattgttcaaaagtgtgctttatgttaaacgacagagactaatgctg 534  
 DB |||||||  
 QY 181 catgcccgttctgctgtaatacagaaggacacatcttgaggctgagatctcagaactgt 240  
 DB |||||||  
 QY 535 catgcccgttctgctgtaatacagaaggacacatcttgaggctgagatctcagaactgt 594  
 DB |||||||  
 QY 241 tctctggaggacccctgtgctcaacttcatcagagacatcactgtcatatagactgt 300  
 DB |||||||  
 QY 595 tctctggaggacccctgtgctcaacttcatcagagacatcactgtcatatagactgt 654  
 DB |||||||  
 QY 301 caagcaaacccctcaaaagtgtgacttggccaacaccttcgttgcttactcagctcag 360  
 DB |||||||  
 QY 655 caagcaaacccctcaaaagtgtgacttggccaacaccttcgttgcttactcagctcag 714  
 DB |||||||  
 QY 361 actcgtactgctcacaacatgctgaactgtcttgaggagataatgcttggaatactatc 420  
 DB |||||||  
 QY 715 actcgtactgctcacaacatgctgaactgtcttgaggagataatgcttggaatactatc 774  
 DB |||||||  
 QY 421 acctcttatagacaacaaatctgtcaaggagcaaaagaaaccttgcataaactcgg 480  
 DB |||||||  
 QY 775 acctcttatagacaacaaatctgtcaaggagcaaaagaaaccttgcataaactcgg 834  
 DB |||||||  
 QY 481 gaccagaagaatgttctctgagaaatgagatctgtgtaccctatggtccaggtctttgag 540  
 DB |||||||  
 QY 835 gaccagaagaatgttctctgagaaatgagatctgtgtaccctatggtccaggtctttgag 894  
 DB |||||||  
 QY 541 tgtgttctgctgagttcttcacatgagatacaagtgtatgagcgagggctgcttcactg 600  
 DB |||||||  
 QY 895 tgtgttctgctgagttcttcacatgagatacaagtgtatgagcgagggctgcttcactg 954  
 DB |||||||  
 QY 601 ctatgtctctcgagatctctggagacacacatctatccgtctccatctgcttgggag 660  
 DB |||||||  
 QY 955 ctatgtctctcgagatctctggagacacacatctatccgtctccatctgcttgggag 1014  
 DB |||||||  
 QY 661 acccagcgccgaaagcgaagactca 687  
 DB |||||||  
 QY 1015 acccagcgccgaaagcgaagactca 1041  
 DB |||||||

## RESULT 11

AAAX28432

ID AAX28432 standard; DNA; 932 BP.

AC AAX28432;

XX 22-JUN-1999 (first entry)

XX EGF-like homologue PRO240 coding sequence.

DE EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX EGF-like homologue PRO240 coding sequence.

XX WO9914327-A2.  
 PN 25-MAR-1999.  
 XX 10-SEP-1998; 98WO-US18824.  
 XX 25-NOV-1997; 97US-0066840.  
 PR 17-SEP-1997; 97US-0059114.  
 PR 18-SEP-1997; 97US-0059117.  
 PR 15-OCT-1997; 97US-0059263.  
 PR 17-OCT-1997; 97US-0062125.  
 PR 17-OCT-1997; 97US-0062285.  
 PR 24-OCT-1997; 97US-0062287.  
 PR 29-OCT-1997; 97US-0062816.  
 PR 29-OCT-1997; 97US-0063704.  
 XX (GETH ) GENENTECH INC.  
 PA Botstein D, Goddard A, Gurney A, Hillian K, Lawrence DA;  
 PI Roy M, Wood WI;  
 XX WPI: 1999-229532/19.  
 DR P-PSDB; AAY05282.  
 XX Antibodies against specific proteins overexpressed in tumours  
 PT Example 1; Fig 11; 130pp; English.

XX This sequence encodes the EGF-like homologue PRO240.  
 CC The invention relates to antibodies (Ab) that bind to any of the  
 CC polypeptides (I) designated PRO187; PRO533; PRO214; PRO240; PRO211;  
 CC PRO230; PRO261; PRO246 or EBAF-2. The Ab, or other agents that inhibit  
 CC expression and/or activity of (I) are used: (i) to inhibit growth of  
 CC tumours; and (ii) as diagnostic/prognostic reagents for detection or  
 CC quantification of (I) in cells or tissues, by standard immunoassays, with  
 CC overexpression being indicative of cancer. For therapeutic use, the Ab  
 CC may be conjugated to a toxin, chemotherapeutic agent or radioisotope.  
 CC Genes expressing (I), many of which are growth factor homologies, are  
 CC overexpressed in some cases of cancer.

SQ Sequence 932 BP; 239 A; 241 C; 212 G; 237 T; 3 other;

Query Match 99.7%; Score 685; DB 20; Length 932;  
 Best Local Similarity 99.7%; Pred. No. 4.8e-210;  
 Matches 685; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atggcgctcaagcgccggtgagttcttcagaccctgtgcttggtgctgcgcctgtc 60  
 DB |||||||  
 QY 12 atggcgctcaagcgccggtgagttcttcagaccctgtgcttggtgctgcgcctgtc 71  
 DB |||||||  
 QY 61 ctgcgtctgggagcgcttgaaaggctctgagcgtaccgagataatgccaatgtccagg 120  
 DB |||||||  
 QY 72 ctgcgtctgggagcgcttgaaaggctctgagcgtaccgagataatgccaatgtccagg 131  
 DB |||||||  
 QY 121 agcgtgcaaaattgttcaaaagtgtgctttatgttaaacgacagagactaatgctg 180  
 DB |||||||  
 QY 132 agcgtgcaaaattgttcaaaagtgtgctttatgttaaacgacagagactaatgctg 191  
 DB |||||||  
 QY 181 catgcccgttctgctgtaatacagaaggacacacatcttgaggctgagatctccagaactgt 240  
 DB |||||||  
 QY 192 catgcccgttctgctgtaatacagaaggacacacatcttgaggctgagatctccagaactgt 251  
 DB |||||||  
 QY 241 tctctggaggacccctgtgctcaaaacttcatcagaacatcactatgcatatagactgt 300  
 DB |||||||  
 QY 252 tctctggaggacccctgtgctcaaaacttcatcagaacatcactatgcatatagactgt 311  
 DB |||||||  
 QY 301 caagcaaacccctcaaaagtgtgacttggccaacaccttcgttgcttactcagctcag 360  
 DB |||||||  
 QY 312 caagcaaacccctcaaaagtgtgacttggccaacaccttcgttgcttactcagctcag 371  
 DB |||||||  
 QY 361 actctgatactgccaacaacatgtcaactgtcttgaggagataatgcttgaatactatc 420  
 DB |||||||

```

|||||
Db 372 accctgatactgcacaacatgtaactgctccggaagaattaaagccggaatactacc 431
OY 421 accctatataagacaacaaatctgtaagggaacaaagaacctgtgacataaactggg 480
Db 432 acccttatataagacaacaaatctgtaagggaacaaagaacctgtgacataaactggg 491
OY 481 gaccagaanaatgtgtccctggaatgatactgtgtacctaagatgctccaggtctttggag 540
Db 492 gaccagaanaatgtgtccctggaatgatactgtgtacctaagatgctccaggtctttggag 551
OY 541 tgggtgtgctgatagtttctggaacacacactctacgtctccatcttctgttggag 600
Db 552 tgggtgtgctgatagtttctggaacacacactctacgtctccatcttctgttggag 611
OY 601 ctatgtctctggaatctggaagcaccacactctacgtctccatcttctgttggag 660
Db 612 ctatgtctctggaatctggaagcaccacactctacgtctccatcttctgttggag 671
OY 661 acccagcgccgaaagcagaacttca 687
Db 672 acccagcgccgaaagcagaacttca 698

RESULT 12
AAC00163
ID AAC00163 standard; cDNA; 459 BP.
XX
AC AAC00163;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 161.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
WP1; 2000-500381/45.
XX
DR P-PSDB; AAG00157.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PI obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 161; 71bp + CD-ROM; English.
XX
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion

```

```

CC vectors.
XX
SQ Sequence 459 BP; 114 A; 134 C; 109 G; 102 T; 0 other;

Query Match          64.5%; Score 443; DB 21; Length 459;
Best Local Similarity 100.0%; Pred. No. 2,2e-132;
Matches 443; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 atgagcgtctacagcgccgggtgtagtcttacgacctgtgtgacctgtgacctgtgtc 60
Db 17 atgagcgtctacagcgccgggtgtagtcttacgacctgtgtgacctgtgtgacctgtgtc 76
OY 61 ctgagctctgagcggtggaagggctctgagcgctacccgagatatagacccaatgtccagg 120
Db 77 ctgagctctgagcggtggaagggctctgagcgctacccgagatatagacccaatgtccagg 136
OY 121 agcgtgcaaatctgttcaaaagtggccttattgtttaaagcacagagagctaatgtctg 180
Db 137 agcgtgcaaatctgttcaaaagtggccttattgtttaaagcacagagagctaatgtctg 196
OY 181 catgcccgttctgtgctgtaatacagaagggcacacatcttggggctgtgattccagaactgt 240
Db 197 catgcccgttctgtgctgtaatacagaagggcacacatcttggggctgtgattccagaactgt 256
OY 241 tctctgagagaccctgtgctcaacttcaacacattcaacgacacatacactgtatcataagactgt 300
Db 257 tctctgagagaccctgtgctcaacttcaacacattcaacgacacatacactgtatcataagactgt 316
OY 301 caagcaaacccctcaaaagtgtactgtgccaacacaccttcgtgtgcttactaactcag 360
Db 317 caagcaaacccctcaaaagtgtactgtgccaacacaccttcgtgtgcttactaactcag 376
OY 361 acctgtatactgccaacatgtcaactgtctctgaggaatatagacctgatactatc 420
Db 377 acctgtatactgccaacatgtcaactgtctctgaggaatatagacctgatactatc 436
OY 421 acctatataagacaacaaat 443
Db 437 acctatataagacaacaaat 459

RESULT 13
AAH12571/C
ID AAH12571 standard; cDNA; 462 BP.
XX
AC AAH12571;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (3'-primer) SEQ ID NO:9406.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX

```



DR WPI: 2001-318749/34.  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs.  
 XX  
 XX Claim 3: SEQ ID 9406; 2537bp + CD ROM; English.  
 CC  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 PS Sequence 462 BP; 132 A; 108 C; 94 G; 112 T; 16 other;  
 SQ  
 Query Match 25.2%; Score 173.2; DB 22; Length 462;  
 Best Local Similarity 88.6%; Pred. No. 1.3e-45;  
 Matches 225; Conservative 0; Mismatches 25; Indels 4; Gaps 4;  
 QY 436 aacaaatctgtcaaggcaaaagaaactgtacatacaactgtggagccca-gaatgtg 494  
 DB 462 AACCAAAATTTTNAAGGGCAAAAGAGCTTNGCANANAMATGGGAGCCAGGAATGTC 403  
 QY 495 tccgtgagaatgagatctgtgtactgtgtgtcaggtctt-gcagtggtgtgtgtgtg 553  
 DB 402 TCCGTGAGAAATGATTTGTCGACGATGTCGAGTNTTTCGCACTGTGTGTGTGTGCA 343  
 QY 554 atggttccatgatacaagtgatgagcagagctgtcttctactgtatattctctg 613  
 DB 342 TGGTTTCCATGAGNACAAG-GTATGCGCCAGGCTTGTCTCA-TGNTTATGTTNTTCG 285  
 QY 614 ggattctggagcccaactctatcgtctcattctgtcttggcagccagcgccga 673  
 DB 284 GGATTNTGAGGACCACTCTATCCGTTCATTTTGTGGGAGACCACGCGCGAA 225  
 QY 674 aagccaagacttca 687  
 DB 224 AAGCCAAGACTTCA 211  
 RESULT 14  
 AAX28449  
 ID AAX28449 standard; DNA; 50 BP.  
 XX AAX28449;  
 AC  
 XX 22-JUN-1999 (first entry)  
 DT  
 XX BEF-like/Fgf-8 homologue coding sequence probe SEQ ID No 67.  
 DE  
 XX Antibody; PRO187; PRO533; PRO214; PRO240; PRO211; PRO230; PRO261; PRO246;  
 KW EBARF-2; Inhibitor; tumour growth; cancer; EGF-like homologue; probe;  
 KW FGF-8 homologue; ss.

XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 XX WO914327-A2.  
 PN  
 XX  
 XX 25-MAR-1999.  
 PD  
 XX 10-SEP-1998; 98WO-US18824.  
 PF  
 XX 25-NOV-1997; 97US-0066840.  
 PR 17-SEP-1997; 97US-0059114.  
 PR 17-SEP-1997; 97US-0059117.  
 PR 18-SEP-1997; 97US-0059263.  
 PR 15-OCT-1997; 97US-0062125.  
 PR 17-OCT-1997; 97US-0062285.  
 PR 17-OCT-1997; 97US-0062287.  
 PR 24-OCT-1997; 97US-0062816.  
 PR 29-OCT-1997; 97US-0063704.  
 XX  
 PA (GENE) GENENTECH INC.  
 XX  
 XX Botstein D, Goddard A, Gurney A, Hillan K, Lawrence DA;  
 PI Roy M, Wood WI;  
 PI  
 XX WPI: 1999-229532/19.  
 PT  
 XX Antibodies against specific proteins overexpressed in tumours  
 PS  
 XX Example 1; Page 43; 130pp; English.  
 CC  
 CC This sequence represents a probe used to isolate DNA encoding a  
 CC protein recognised by the antibodies of the invention.  
 CC The invention relates to antibodies (Ab) that bind to any of the  
 CC polypeptides (I) designated PRO187; PRO533; PRO214; PRO240; PRO211;  
 CC PRO230; PRO261; PRO246 or EBARF-2. The Ab, or other agents that inhibit  
 CC expression and/or activity of (I) are used: (i) to inhibit growth of  
 CC tumours; and (ii) as diagnostic/prognostic reagents for detection or  
 CC quantification of (I) in cells or tissues, by standard immunoassays, with  
 CC overexpression being indicative of cancer. For therapeutic use, the Ab  
 CC may be conjugated to a toxin, chemotherapeutic agent or radioisotope.  
 CC Genes expressing (I), many of which are growth factor homologues, are  
 CC overexpressed in some cases of cancer.  
 CC  
 SQ Sequence 50 BP; 12 A; 10 C; 15 G; 13 T; 0 other;  
 Query Match 7.3%; Score 50; DB 20; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-06;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 480 ggaccagaagaatgtctcctgagaatgagatctgtgtactgtatgtgtcag 529  
 DB 1 ggaccagaagaatgtctcctgagaatgagatctgtgtactgtatgtgtcag 50  
 RESULT 15  
 AAC58413  
 ID AAC58413 standard; DNA; 50 BP.  
 XX AAC58413;  
 AC  
 XX 29-JAN-2001 (first entry)  
 DT  
 XX Human PRO240 (UNC214) hybridisation probe SEQ ID No:29.  
 DE  
 XX Human; immune related disease; diagnosis; antiinflammatory; cardiant;  
 KW dermatological; antidiabetic; antihemetic; immunosuppressive;  
 KW haemostatic; antihypertensive; antidiabetic; neuroprotective;  
 KW antianemic; hepatotropic; virucide; antiparasitic; antiallergic;  
 KW antiasthmatic; systemic lupus erythematosus; Rheumatoid arthritis;  
 KW osteoarthritis; spondyloarthropathy; systemic sclerosis; sarcoidosis;  
 KW Idiopathic inflammatory myopathy; Sjogren's syndrome; thyroiditis;



KW systemic vasculitis; autoimmune haemolytic anaemia; diabetes mellitus;  
 KW autoimmune thrombocytopenia; immune-mediated renal disease;  
 KW demyelinating disease; hepatobiliary disease; Whipple's disease;  
 KW inflammatory bowel disease; gluten-sensitive enteropathy; hybridisation;  
 KW autoimmune disease; immune-mediated skin disease; allergic disease;  
 KW immunological disease; transplantation associated disease; PCR primer;  
 \*KW graft rejection; graft-versus-host-disease; probe; ss.  
 XX  
 OS Homo sapiens.  
 PN WO200053758-A2.  
 XX  
 PD 14-SEP-2000.  
 XX  
 PF 02-MAR-2000; 2000WO-US05841.  
 XX  
 PR 08-MAR-1999; 99WO-US05028.  
 PR 10-MAR-1999; 99US-0123618.  
 PR 12-MAR-1999; 99US-0123957.  
 PR 23-MAR-1999; 99US-0125775.  
 PR 12-APR-1999; 99US-0128849.  
 PR 20-APR-1999; 99WO-US08615.  
 PR 28-APR-1999; 99US-0131445.  
 PR 04-MAY-1999; 99US-0132371.  
 PR 14-MAY-1999; 99US-0134287.  
 PR 02-JUN-1999; 99WO-US12252.  
 PR 23-JUN-1999; 99US-0141037.  
 PR 20-JUL-1999; 99US-0144758.  
 PR 26-JUL-1999; 99US-0145698.  
 PR 28-JUL-1999; 99US-0146222.  
 PR 01-SEP-1999; 99WO-US20111.  
 PR 08-SEP-1999; 99WO-US20594.  
 PR 13-SEP-1999; 99WO-US20944.  
 PR 15-SEP-1999; 99WO-US21090.  
 PR 15-SEP-1999; 99WO-US21547.  
 PR 05-OCT-1999; 99WO-US23089.  
 PR 29-OCT-1999; 99US-0162506.  
 PR 29-NOV-1999; 99WO-US28214.  
 PR 30-NOV-1999; 99WO-US28313.  
 PR 30-NOV-1999; 99WO-US28409.  
 PR 01-DEC-1999; 99WO-US28301.  
 PR 01-DEC-1999; 99WO-US28634.  
 PR 02-DEC-1999; 99WO-US28551.  
 PR 02-DEC-1999; 99WO-US28564.  
 PR 02-DEC-1999; 99WO-US28565.  
 PR 16-DEC-1999; 99WO-US30095.  
 PR 20-DEC-1999; 99WO-US30999.  
 PR 30-DEC-1999; 99WO-US31274.  
 PR 05-JAN-2000; 2000WO-US00219.  
 PR 06-JAN-2000; 2000WO-US00277.  
 PR 06-JAN-2000; 2000WO-US00376.  
 PR 11-FEB-2000; 2000WO-US03565.  
 PR 18-FEB-2000; 2000WO-US04341.  
 PR 18-FEB-2000; 2000WO-US04342.  
 PR 22-FEB-2000; 2000WO-US04414.  
 XX  
 PA (GENTECH ) GENENTECH INC.  
 XX  
 PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hebert C, Henzel W;  
 PI Kabakoff RC, Lu Y, Pan Y, Pennica D, Shelton DL, Smith V;  
 PI Stewart TA, Tumas D, Watanabe CK, Wood WI, Yan M;  
 XX  
 DR WPI; 2000-572271/53.  
 XX  
 PT Sixty four PRO polypeptides, useful in the diagnosis and treatment of  
 PT immune related disorders, e.g. systemic lupus erythematosus, rheumatoid  
 PT arthritis, osteoarthritis, thyroiditis and diabetes mellitus -  
 XX  
 PS Example 1; Page 94; 309pp; English.  
 CC The present invention describes sixty four human PRO proteins which can  
 CC be used in the treatment of immune related diseases. The human PRO  
 CC proteins, anti-PRO antibodies, agonists and antagonists are useful for

CC treating and diagnosing immune related disorders. The disorders are  
 CC selected from systemic lupus erythematosus, rheumatoid arthritis,  
 CC osteoarthritis, juvenile chronic arthritis, spondyloarthropathies,  
 CC systemic sclerosis, idiopathic inflammatory myopathies, Sjogren's  
 CC syndrome, systemic vasculitis, sarcoidosis, autoimmune haemolytic  
 CC anaemia, autoimmune thrombocytopenia, thyroiditis, diabetes mellitus,  
 CC immune-mediated renal disease, demyelinating diseases of the central  
 CC and peripheral nervous systems, hepatobiliary diseases, inflammatory  
 CC bowel disease, gluten-sensitive enteropathy and Whipple's disease,  
 CC autoimmune or immune-mediated skin diseases, allergic diseases,  
 CC immunological diseases of the lung, and transplantation associated  
 CC diseases including graft rejection and graft-versus-host-disease.  
 CC AAC58397 to AAC58578 represent PCR primers and hybridisation probes used  
 CC in the isolation of human PRO sequences. AAC58579 to AAC58642 and  
 CC AAB33414 to AAB33477 represent human PRO polynucleotide and protein  
 CC sequences given in the exemplification of the present invention.  
 XX  
 SQ Sequence 50 BP; 12 A; 10 C; 15 G; 13 T; 0 other;

Qy 480 ggaccagaatgtgtctgaagaatgatctgtgtactgtatgtccag 529  
 |||||||  
 Db 1 ggaccagaatgtgtctgaagaatgatctgtgtactgtatgtccag 50

Search completed: March 28, 2002, 07:33:14  
 Job time: 10224 sec

Thu Mar 28 09:21:27 2002

us-09-726-348-1\_copy\_5\_691.rng

---

Page 14

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 02:13:42 ; Search time 88.1 Seconds  
(without alignments)  
1766.066 Million cell updates/sec

Title: US-09-726-348-1\_COPY\_5\_691

Perfect score: 687  
Sequence: 1 atgagcctcaagccgcggg.....gccgaagcaagactca 687

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32.6	4.7	5438	US-08-456-200B-5	Sequence 5, Appli
2	32.6	4.7	7175	US-08-455-543A-8	Sequence 8, Appli
3	32.6	4.7	7175	US-08-193-078B-8	Sequence 8, Appli
4	32.6	4.7	7175	US-08-223-305C-8	Sequence 8, Appli
5	32.6	4.7	7175	US-08-149-097D-8	Sequence 8, Appli
6	32.6	4.7	7175	US-08-949-386-8	Sequence 8, Appli
7	32.6	4.7	7175	US-08-450-562-8	Sequence 8, Appli
8	32.6	4.7	7266	US-08-713-118-1	Sequence 1, Appli
9	32.6	4.7	7266	US-09-452-007-1	Sequence 1, Appli
10	32.6	4.7	7362	US-08-455-543A-7	Sequence 7, Appli
11	32.6	4.7	7362	US-08-193-078B-7	Sequence 7, Appli
12	32.6	4.7	7362	US-08-223-305C-7	Sequence 7, Appli
13	32.6	4.7	7362	US-08-149-097D-7	Sequence 7, Appli
14	32.6	4.7	7362	US-08-949-386-7	Sequence 7, Appli
15	32.6	4.7	7362	US-08-450-562-7	Sequence 7, Appli
16	32.6	4.7	7362	US-08-455-543A-7	Sequence 7, Appli
17	32.6	4.7	7362	US-08-193-078B-7	Sequence 7, Appli
18	32.6	4.7	7362	US-08-223-305C-7	Sequence 7, Appli
19	32.6	4.7	7362	US-08-149-097D-7	Sequence 7, Appli
20	32.6	4.7	7362	US-08-949-386-7	Sequence 7, Appli
21	32.6	4.7	7362	US-08-450-562-7	Sequence 7, Appli
22	32.6	4.7	7362	US-08-455-543A-7	Sequence 7, Appli
23	32.6	4.7	7362	US-08-193-078B-7	Sequence 7, Appli
24	32.6	4.7	7362	US-08-223-305C-7	Sequence 7, Appli
25	32.6	4.7	7362	US-08-149-097D-7	Sequence 7, Appli
26	32.6	4.7	7362	US-08-949-386-7	Sequence 7, Appli
27	32.6	4.7	7362	US-08-450-562-7	Sequence 7, Appli

28	29.2	4.3	2745	2	US-08-658-665-62	Sequence 62, Appli
29	29.2	4.3	2745	4	US-08-796-101-26	Sequence 26, Appli
30	29.2	4.3	2745	4	US-09-085-273-62	Sequence 62, Appli
31	29.2	4.2	1644	3	US-08-948-564-5	Sequence 5, Appli
32	28.8	4.2	4666	4	US-08-821-278A-16	Sequence 16, Appli
33	28.8	4.2	2775	1	US-08-481-130-25	Sequence 25, Appli
34	28.8	4.2	2775	1	US-08-656-984A-25	Sequence 25, Appli
35	28.8	4.2	2775	1	US-08-485-604-25	Sequence 25, Appli
36	28.8	4.2	2775	2	US-08-487-595-25	Sequence 25, Appli
37	28.8	4.2	2927	1	US-08-481-130-27	Sequence 27, Appli
38	28.8	4.2	2927	1	US-08-656-984A-27	Sequence 27, Appli
39	28.8	4.2	2927	1	US-08-485-604-27	Sequence 27, Appli
40	28.8	4.2	2927	2	US-08-487-595-27	Sequence 27, Appli
41	28.8	4.2	3024	1	US-08-149-100-1	Sequence 1, Appli
42	28.8	4.2	46899	1	US-08-471-119A-1	Sequence 1, Appli
43	28.6	4.2	2582	1	US-08-514-014-3	Sequence 3, Appli
44	28.6	4.2	2582	2	US-08-833-823-3	Sequence 3, Appli
45	28.6	4.2	2775	1	US-08-730-771-1	Sequence 1, Appli

# ALIGNMENTS

RESULT 1  
US-08-456-200B-5  
Sequence 5, Application US/08456200B  
Patent No. 6229000  
GENERAL INFORMATION:  
APPLICANT: Franz, Jurgen; Weingartner, Bernhard;  
INVENTOR: Unterbeck, Axel; Rae, Peter  
TITLE OF INVENTION: TISSUE-SPECIFIC HUMAN NEURONAL  
TITLE OF INVENTION: CALCIUM CHANNEL SUB-TYPES AND  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESS: SPRUNG HORN KRAMER & WOODS  
STREET: 660 White Plains Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 1059. 1144  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB  
MEDIUM TYPE: storage  
COMPUTER: NEC Powermate SX/20  
OPERATING SYSTEM: DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,200B  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/094,712  
FILING DATE: 19-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/858,278  
FILING DATE: 26-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/064,778  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 41 10 785  
FILING DATE: 04-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurt G. Briscoe  
REGISTRATION NUMBER: 33,141  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 332-1700  
TELEFAX: (914) 332-1844  
TELEX:  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:



```

Query Match          4.7%:  Score 33.6:  DB 2,  Length 7175:
Best Local Similarity 49.1%:  Pred. No. 1.7:
Matches      86:  Conservative      0:  Mismatches      89:  Indels      0:  Gaps      0:

Db      157  aaagacacgacgacatcgtcgtatgcgcgttctctgcctcaatcagaagggcacacc 216
          |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
Db      5745  AAACACCACGAGACCAAGTATGACGAGAGGTCTCTGGAGGSCCTCTCCACATGGGTCTCTG 5804
          |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||

Db      217  ttgggctgagatcctccagaactgttctctgtagagacccctggtccaaacttcatacgaca 276
          |||||  ||||  ||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      5805  TCCCTGTTCACCCCTCTCTGAGGCGCACCCCTGGAGGACAGACACCGGCTTGTCTCCGAGGA 5864
          |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||

Db      277  catcacactgcatatagacctgcaagaacccccctcaagaagtactggaca 331
          |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||

Db      5865  GCCCGGGTTTCTCTTCGACAGAGAGTTCCACCTCCCTTAGCANATGCGGGGGCCA 5919
          |||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||

RESULT      4
US-08-223-305C-8
: Sequence 8, Application US/08223305C
: Patent No. 5851824
: GENERAL INFORMATION:
: APPLICANT: Harpold, Michael
: APPLICANT: Ellis, Steven
: APPLICANT: Williams, Mark
: APPLICANT: Feldman, Daniel
: APPLICANT: McCue, Ann
: APPLICANT: Brenner, Robert
: TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
: METHODS
: NUMBER OF SEQUENCES: 57
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Brown, Martin, Haller & McClain
: STREET: 1660 Union Street
: CITY: San Diego
: STATE: California

```

Query Match	4.7%	Score 32.6	DB 2	Length 7175
Best Local Similarity	49.1%	Pred. No. 1.7		
Matches	86	Conservative	0	Mismatches 89
				Indels 0
				Gaps 0
QY 157	aaagacacgagagctatgctgcatccgcttctgctcattacagaagcgaccalc	216		
DB 5745	AAACCACCACGAGACCAGATTGCAGAGGCTCTTGAGGSCCTTCCCAATGGGTCTGTG	5804		
QY 217	ttagggctgtagatctccagacgcttctctctgagagacccctgctccaactttcatcaagca	276		
DB 5805	TGCCCTTCCACCCCTCTGAAGGCCACCCCTGGAGACAGACACCGGCTGTGCTCCGAGA	5864		
QY 277	cataccactgcatcatagactgcgaagcaaaacccctcaaaagtgaacttgcca	331		
DB 5865	GCCCGGGTTTCTCTTCGACAGAAGTTCCACTCCTCAGCAATGGCGGGGCCA	5919		

RESULT 5  
US-08-149-097D-8  
Sequence 8, Application US/08149097D  
Patent No. 5874236  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: Feldman, Daniel  
APPLICANT: McQue, Ann  
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
TITLE OF INVENTION: METHODS  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/149,097D  
FILING DATE: 05-NOV-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/105,536  
FILING DATE: 11-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US92/06903  
FILING DATE: 14-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/914,231  
FILING DATE: 13-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/868,354  
FILING DATE: 10-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/745,206  
FILING DATE: 15-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/620,250  
FILING DATE: 30-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/482,384  
FILING DATE: 20-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/603,751  
FILING DATE: 04-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US89/01408  
FILING DATE: 04-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/176,899  
FILING DATE: 04-APR-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-55038  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 238-0999  
TELEFAX: (619) 238-0062  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7175 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 144..6857  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..143  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 6855..7175  
US-08-149-097D-8  
Query Match 4.7%; Score 32.6; DB 2; Length 7175;  
Best Local Similarity 49.1%; Pred. No. 1.7;  
Matches 86; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 157 aaagacagacagagcttaagctgcatgcccgtgtgcttgatataagaagggacaccac 216  
DB 5745 AAAACCAACAGAGACAGATGACGAGGCTCTGAGGCTCTCCCAATGGCTCTGTG 5804  
QY 217 ttggagctgagatccagactgtctctggaagaccctggtccaaacttcataagca 276  
DB 5805 TCCCTGTTCCACCCCTCTGAAGGCCACCCCTGGACACACAGCGGCTGTCTCCAGCA 5864  
QY 277 catcacctgtcatcatgacctgcaagaacacccctcaaaagtgacttgcca 331  
DB 5865 GCCCGGGTTTCCCTCGACAGAGAGTTCACCTCCCTCGAATGGCGGGGCCA 5919

RESULT 6  
US-08-949-386-8  
Sequence 8, Application US/08949386  
Patent No. 6090623  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: McQue, Ann  
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
TITLE OF INVENTION: METHODS  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/949,386  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,012  
FILING DATE: 11-AUG-1994  
APPLICATION NUMBER: 08/149,097  
FILING DATE: 5-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/105,536  
FILING DATE: 11-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 519808  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 238-0999  
TELEFAX: (619) 238-0062  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7175 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 144..6857  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..143  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 6855..7175  
US-08-949-386-8

Query Match 4.7%; Score 32.6; DB 3; Length 7175;  
Best Local Similarity 49.1%; Pred. No. 1.7;  
Matches 86; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 157 aaagacacacagagtaagtctgcatgcccgtgtctgctgaatcagaagggcaccatc 216  
Db 5745 AAACACACACAGAGATGACGAGGCTCTCGAGGCTCTCCAGATGGGCTCTGTG 5804  
Qy 217 ttggagctgcatcctcagactgtctctggaagaccctgttccaaatttcacagca 276  
Db 5805 TCCCTGTCCACCCCTGGAAGGCCCTCGACACACACGCGGTGTGTCGCGAGA 5864  
Qy 277 catcaactgtcatcagactgcaagcgaaccctcaaaagttaacttgccca 331  
Db 5865 GCCCGGGTTTCTTGACAGAGAGTCCACCTCCTCGAATGGCGGGGCCA 5919

RESULT 7  
US-08-450-562-8  
Sequence 8, Application US/08450562  
Patent No. 6096514  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: McCue, Ann  
APPLICANT: Gillespie, Alison  
APPLICANT: Feldman, Daniel  
APPLICANT: Brenner, Robert  
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
TITLE OF INVENTION: METHODS  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: California  
COUNTRY: US  
ZIP: 92101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,562  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/404,950  
FILING DATE: 13-MAR-1995  
APPLICATION NUMBER: 08/336,257

FILING DATE: 7-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/314,083  
FILING DATE: 28-SEPT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/311,363  
FILING DATE: 23-SEPT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/290,012  
FILING DATE: 11-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/223,305  
FILING DATE: 4-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/193,078  
FILING DATE: 07-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/149,097  
FILING DATE: 5-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/105,536  
FILING DATE: 11-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/914,231  
FILING DATE: 13-JULY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/868,354  
FILING DATE: 10-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/06903  
FILING DATE: 14-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/745,206  
FILING DATE: 15-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/620,250  
FILING DATE: 30-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/603,751  
FILING DATE: 08-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/482,384  
FILING DATE: 02-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US89/01408  
FILING DATE: 04-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/176,899  
FILING DATE: 04-APR-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-519812  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 238-0062  
TELEFAX: (619) 238-0062  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7175 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 144..6857  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..143  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 6855..7175  
US-08-450-562-8







```

RESULT 13
US-08-149-097D-7
: Sequence 7, Application US/08149097D
: Patent No. 5874236
:
: GENERAL INFORMATION:
:
: APPLICANT: Harpold, Michael
: APPLICANT: Ellis, Steven
: APPLICANT: Williams, Mark
: APPLICANT: Feldman, Daniel
: APPLICANT: McCue, Ann
: APPLICANT: Brenner, Robert
: TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
: TITLE OF INVENTION: METHODS
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Brown, Mattin, Haller & McClain
:
: STREET: 1660 Union Street
: CITY: San Diego
: STATE: California
: COUNTRY: USA

```

```

ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,097D
FILING DATE: 05-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US92/06903
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/868,354
FILING DATE: 10-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/745,206
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/620,250
FILING DATE: 30-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/482,384
FILING DATE: 20-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,751
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US89/01408
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/176,899
FILING DATE: 04-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-55038
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7362 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 144..7163
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..143
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 7161..7362

```

Query	March	Similarity	49.1%	Score	32.6	DB	2	Length	7362
Best	Local	Similarity	49.1%	Pred.	No. 1.7				
Matches	86	Conservative	0	Mismatches	89	Indels	0	Gaps	0
QY	157	aaacagcacagagagctaatctgcgtatccgtctgtctgtcgttaacaaagaagcacaccalc	216						
DB	5745	AAACCCACACAGACAGATGTCAGGAGGCTCTCTGAGGCGCTCTCCCAATAGCTCTCTGTG	5804						

OY	217	ttggagctcgtgaatccccaagaactgtcttccttagaggaacccctggtcccaacttcaatcagsga	276
Db	5805	ttccctgttccacccctctgaaagccaccctctgagcagacagacaaagccggctgtgctnccagga	586
OY	217	cataccactgtgcatcatagactctcaagcaaaacccctcaagatgacttgcsgca	331
Db	5865	gcccgcggttttcccttccgacagaaagatgtccacacctcctctcagcaaatggcggggcca	5919

RESULT 14  
US-08-949-386-7

Sequence /, Application 05/06345500  
; Patent No. 6090623  
; Patent No. 6090623  
GENERAL INFORMATION:

GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPOINTMENT: 11/18/2000

APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: McQue, Ann

APPLICANT: MCCUE, ANN  
APPLICANT: GILLESPIE, ALISCO  
TITLE OF INVENTION: HUMAN O

[illegible]

NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown Martin, Haller & McClain

ADDRESS: (Down) main  
STREET: 1660 Union Street  
CITY: San Diego

CITY: San Diego  
STATE: California  
COUNTRY: U.S.

COMPUTER READ

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patentIn Release #1.0

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/949,386

```

:  
 : FILING DATE:  
 :  
 : CLASSIFICATION:  
 :

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,012

FILING DATE: 11-AUG-1994  
APPLICATION NUMBER: 08/149,097

FILING DATE: 5-NOV-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/105,536  
FILING DATE: 11-AUG-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.

REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 519808

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 238-0999

TELEFAX: (619) 238-0062  
; INFORMATION FOR SEQ ID NO: 7:  
;

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 7362 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
;

```

```

;
;   TOPOLOGY: linear
;
;   MOLECULE TYPE: DNA (genomic)
;

```

```

;
;   FEATURE:
;   NAME/KEY:  CDS
;

```

```

; LOCATION: 144..7163
;
; FEATURE:
;

```

```

; NAME/KEY: 5.01R
; LOCATION: 1..143
;

```

```

;
; FEATURE:
; NAME/KEY: 3'UTR
;

```

LOCATION: /161.../362  
US-08-949-386-7

\_\_\_\_\_

```

1      FILING DATE: 11-AUG-1993
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: 07/614,231
4      FILING DATE: 13-JULY-1992
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: 07/868,354
7      FILING DATE: 10-APR-1992
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: PCT/US92/06903
10     FILING DATE: 14-AUG-1992
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: 07/745,206
13     FILING DATE: 15-AUG-1991
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: 07/620,250
16     FILING DATE: 30-NOV-1990
17     PRIOR APPLICATION DATA:
18     APPLICATION NUMBER: 07/603,751
19     FILING DATE: 08-NOV-1990
20     PRIOR APPLICATION DATA:
21     APPLICATION NUMBER: 07/482,384
22     FILING DATE: 02-FEB-1990
23     PRIOR APPLICATION DATA:
24     APPLICATION NUMBER: PCT/US89/01408
25     FILING DATE: 04-APR-1989
26     PRIOR APPLICATION DATA:
27     APPLICATION NUMBER: 07/176,899
28     FILING DATE: 04-APR-1988
29     ATTORNEY/AGENT INFORMATION:
30     NAME: Selman, Stephanie L.
31     REGISTRATION NUMBER: 33,779
32     REFERENCE/DOCKET NUMBER: 6362-519812
33     TELECOMMUNICATION INFORMATION:
34     TELEPHONE: (619) 238-0999
35     TELEFAX: (619) 238-0062
36     INFORMATION FOR SEQ ID NO: 7:
37     SEQUENCE CHARACTERISTICS:
38     LENGTH: 7362 base pairs
39     TYPE: nucleic acid
40     STRANDEDNESS: double
41     TOPOLOGY: linear
42     MOLECULE TYPE: DNA (genomic)
43     FEATURE:
44     NAME/KEY: CDS
45     LOCATION: 144..7163
46     FEATURE:
47     NAME/KEY: 5'UTR
48     LOCATION: 1..143
49     FEATURE:
50     NAME/KEY: 3'UTR
51     LOCATION: 7161..7362
52     US-08-450-562-7

```

[illegible]

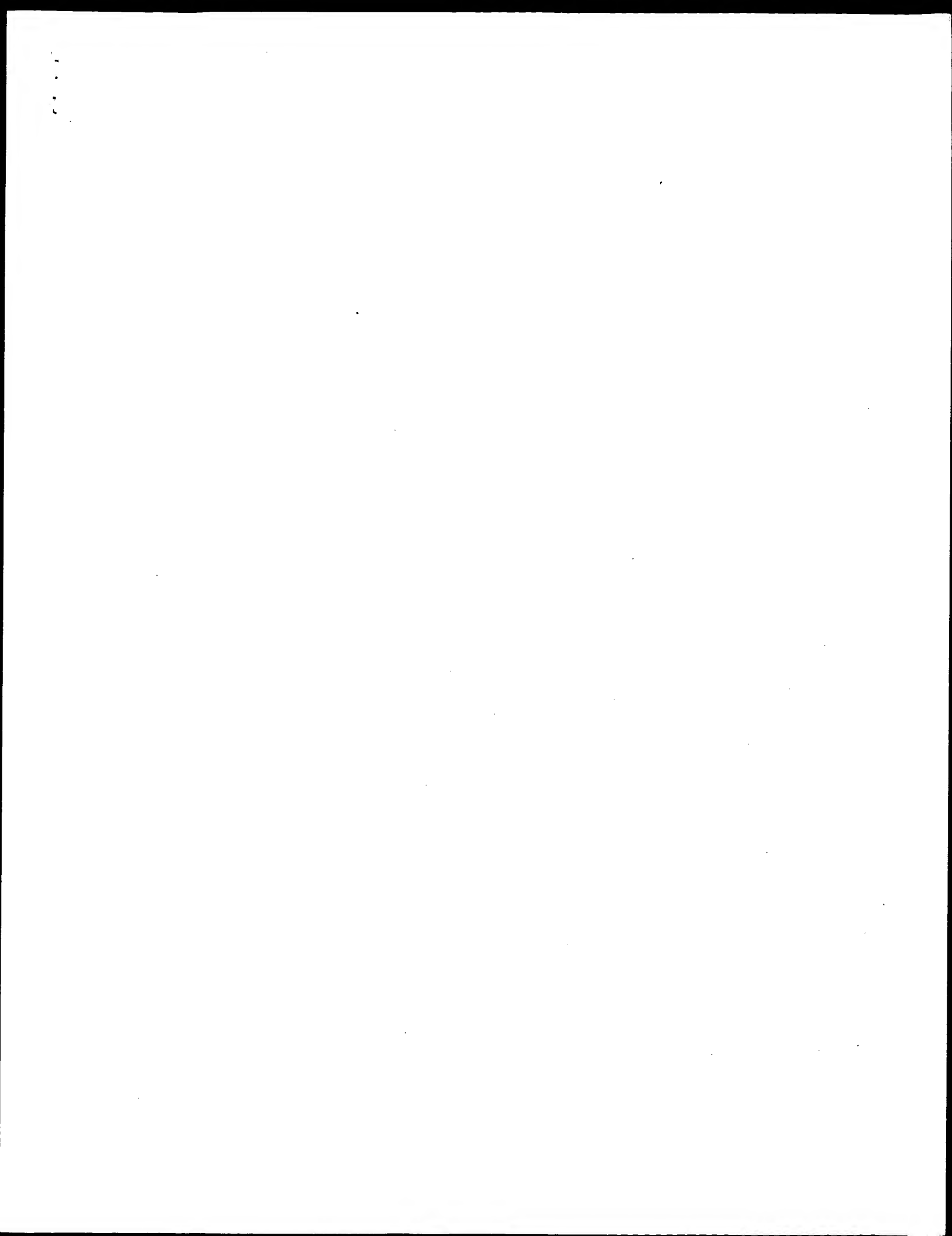
'Thu Mar 28 09:21:28 2002

us-09-726-348-1\_copy\_5\_691.rni

Page 11

Job time: 19016 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:01:29 ; Search time 74.25 Seconds

(without alignments)  
181.588 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_1\_177  
Page: 671

Sequence: 1 MAPHGPGLTTLVPWAALL.....NTGDPENCPCENGSCVPDGP 177

Scoring table:

	BLOSUM62	Consort	Consort
Consort	10	0	0

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

```
PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Query No.	Score	Match	Length	DB	ID	Description
1	114.5	11.8	2907	2	A57378		fibrillin-2 precursor
2	105.5	10.2	2918	2	A54105		fibrillin-2 precursor
3	99.9	10.2	1054	2	B86465		probable protein K
4	95.3	9.8	3002	2	A47221		fibrillin 1 precursor
5	91.9	9.4	1114	2	I50222		deltaE1 - chicken
6	88.5	9.1	2871	2	A55624		fibrillin-1 precursor
7	86.5	8.9	1134	2	T04587		hypothetical protein
8	86.5	8.9	2871	2	A55667		fibrillin 1 - bovin
9	85.5	8.8	389	2	T23167		hypothetical protein
10	82.5	8.5	851	2	B8445		hypothetical protein
11	82	8.4	670	2	I65967		probable receptor
12	82	8.4	932	2	T05325		disintegrin-like m
13	81.5	8.4	639	2	T05325		hypothetical protein
14	81.5	8.4	1011	2	T45718		extensin homolog F
15	81	8.3	432	2	I56934		receptor-kinase 11
16	81	8.3	432	2	A27447		fibrinogen-like protein
17	81	8.3	432	2	C96772		cytotoxic T-lymphocyte
18	80.5	8.3	1130	2	B86479		probable receptor
19	80	8.2	879	2	T02728		hypothetical protein
20	79	8.1	1599	2	T16210		serine/threonine-s
21	78.5	8.1	1003	2	T05898		hypothetical protein
22	78.5	8.1	1450	2	T30273		hypothetical protein
23	78	8.0	979	2	D96574		hypothetical protein
24	77	7.9	766	2	B85440		hypothetical protein
25	77	7.9	767	2	B84594		receptor Kinase-11
26	76.5	7.9	768	2	T17462		probable LRR receptor
27	76.5	7.9	3633	2	T09456		disease resistance
28	76	7.8	3922	2	T23573		intrinsic factor-F
29	75.5	7.8	332	2	T21458		hypothetical protein

30	75.5	7.8	496	2	C96832	hypothetical protein
31	75.5	7.8	254	2	T31687	surface antigen - P
32	75	7.7	886	2	A57172	probable hormone r
33	75	7.7	1199	2	T47442	disease resistance
34	74.5	7.7	855	2	T07015	cf-4A protein - to
35	74	7.6	638	1	KOHUP	plasma kallikrein
36	73.5	7.6	818	2	P95866	hypothetical protein
37	73.5	7.6	1009	2	S61174	hypothetical protein
38	73.5	7.6	2946	2	T15840	hypothetical protein
39	73	7.5	377	2	S77134	hypothetical protein
40	73	7.5	943	2	E84429	probable receptor--
41	73	7.5	979	2	A39792	transcription activi
42	73	7.5	1143	2	B84431	probable receptor
43	73	7.5	1391	2	T20406	hypothetical protein
44	72.5	7.5	195	2	D72254	glycerol uptake op
45	72.5	7.5	1019	1	A56318	enteropeptidase (E

## ALIGNMENTS

RESULT 1  
A57278  
fibrillin-2 precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 11-Jan-2000  
C:Accession: A57278  
R:Zhang, H.; Hu, W.; Ramirez, F.  
J. Cell Biol. 129, 1165-1176, 1995  
A:Title: Developmental expression of fibrillin genes suggests heterogeneity of extracellular matrix  
A:Reference number: A57278; MUID:95263670  
A:Accession: A57278  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-2907 <ZHA>  
A:Cross-references: GB:139790; NID:g762830; PID:AAA74908.1; PID:g762831  
F:1298-1274/Domain: EGF homology <EGF1>  
F:2486-2523/Domain: EGF homology <EGF>

```

Query Match Similarity 11.8%; Score 114.5; DB 2; Length 2907;
Best Local Similarity 23.7%; Pred. No. 0.033;
Matches 47; Conservative 19; Mismatches 69; Indels 63; Gaps 11;

QY      1  MAPPGSGSLITLVWMAALLALGVERALALDETCT--QCPGSOVNISKVAFYCKTTREL 58
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2429  TCPHGGGYATD-----GRDDECKVMSPLCTNSQC---VMTGMSFRCFCVGYTM 2475

QY      59  MLHAFCCINQKGTILGLDNLONGSLEDPPG-PPHQAHT------VIIDLQANPKG 107
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2476  DISGTAC-----VDLDECS-QSPKPCFNICKNTGSGVOCSPGPRVIAQEDGKTK- 2524

QY      108  DL-----ANTFPGTLOQLLLPHQVNCPGGINAWNTITSYDNDQIGQOK 153
           || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2525  DLDPPGCKQKHNCOQLCVNTLGGFT-----CKCPPGFT--TACIDNNCGSQP 2572

QY      154  NLGNATCTCTTMCPCENGSC 171
           || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      2573  SLGCKRGICQNTPGSFSC 2590

RESULT      2
A54105
fibritillin-2 precursor - human
C:Species: Homo sapiens (man)
C:Date: 09-Sep-1994 #sequence revision 09-Sep-1994 #text_change 17-Nov-2000
C:Accession: A54105; S17063; S31101
R:Zhang, H.; Apfelroth, S.D.; Hu, W.; Davis, E.C.; Sanguinetti, C.; Bonadio, J.; Mecha-
J.; Cell Biol. 124, 855-863, 1994
A:Title: Structure and expression of fibritillin-2, a novel microfibrillar component pr
A:Accession: A54105; MUID:94165150

```





R.Madox, B.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W.  
 J. Biol. Chem. 264, 21381-21385, 1989  
 A:Title: Connective tissue microfibrils. Isolation and characterization of three large F  
 A:Reference number: A34198; MUID:90078246  
 A:Accession: A34198  
 A:Molecule type: protein  
 A:Residues: 565-575;1890-1892, 'T', 1894-1900 <MAD>  
 C:Comment: fibrillin is a major component of elastin-associated microfibrils.  
 C:Genetics:  
 A:Gene: GDB:FBN1  
 A:Cross-references: GDB:127115; OMTM:134797; OMTM:154700  
 A:Map position: 15q21.1-15q21.1  
 A:Introns: 2256/1; 2258/1; 2297/1  
 C:Superfamily: unassigned EGF-related proteins; EGF homology  
 C:Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein; N  
 F:1-3002/Product: fibrillin (5'-region exon A splice form) (fragment) #status predicted  
 F:1332-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MAD>  
 F:1332-1367/Domain: EGF homology <EGF>  
 F:1457-1492/Domain: EGF homology <EGF>  
 F:2262-2295/Domain: EGF homology <EGF>

Query Match 9.8%; Score 95.5; DB 2; Length 3002;  
 Best Local Similarity 21.4%; Pred. No. 2.1;  
 Matches 43; Conservative 20; Mismatches 69; Indels 69; Gaps 9;

OY 1 MAPHPSGLTTLVPMALALLALGERALALPEICT--QC---PGSVONLSKVA----- 49  
 Db 2520 LCPHGRGFMN-----GADIDCKVIHDVCRNGECVNDKRSYHCICKTGTPDIT 2569  
 OY 50 -----FYCKTTRELMHARCLNOKGTIIGLDLQNC-SLEDGPNFH 90  
 Db 2570 GTSCVDLNECNOAPKPCNFICKNTE---GSYOCSPKGYIIQEDGRSKDLDECATKQH 2625  
 OY 91 QAHFTVILDLQANPLKGLDANLFRGFTQLQTLILPQHNCGGINAMNTITSYDNOICQ 150  
 Db 2626 NCQFLCV-----NTIGGFT-----CKCPGFTQHH--TSCIDNNECT 2660  
 OY 151 GOKNLCNNTGDEPMCPENGSC 171  
 Db 2661 SDINLCGSKGICQNTPGSFTC 2681

RESULT 5  
 150222  
 deltaE1 - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: 150222  
 R:Funahashi, J.; Sekido, R.; Murai, K.; Kamachi, Y.; Kondoh, H.  
 Development 119, 433-446, 1993  
 A:Title: Delta-crystallin enhancer binding protein delta E1 is a zinc finger-homeodoma  
 A:Reference number: 150222; MUID:94116444  
 A:Accession: 150222  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1114 <FUN>  
 A:Cross-references: GB:D14313; NID:g391631; PIDN:BA03259.1; PID:g391632

Query Match 9.4%; Score 91; DB 2; Length 1114;  
 Best Local Similarity 21.3%; Pred. No. 1.8;  
 Matches 42; Conservative 22; Mismatches 73; Indels 60; Gaps 7;

OY 24 GVERALALPEICTGCP-----GSVONLSKVAFYCKTRELML-LHAACCLNOKGTIIGLD 76  
 Db 389 GYVAVVILPTGLVSPISINLSDIQNLKVAVDGNYLROYLENNHANLASKEDFTISNAS 448  
 OY 77 LQ-----NCSLEDPG-----PNEHQHTVILDLQANPL 105  
 Db 449 IQQAGHLSAISLPLVDODGTIKIITNSLEQPSQLQVVPQNKKHVSPTNSCKNEKL 508  
 OY 106 KGDLA-----NTRFGFTQLQTLILPQHNCGGINAMNTITSYDNOICGOKNLCNNT 159

Db 509 PEDLVYSKKRKNRGEFTNDSTCLCD--DCPGDLNALQELKHT-----ET 552  
 OY 160 GDPMPENGSVCVDPGP 176  
 Db 553 KNPQLPQSSGTFAEKRP 569

RESULT 6  
 A53624  
 fibrillin-1 precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 11-Jan-2000  
 C:Accession: A53624  
 R:Yin, W.; Smiley, E.; Germiller, J.; Sanguinetti, C.; Lawton, T.; Pereira, L.; Ramire  
 J. Biol. Chem. 270, 1798-1806, 1995  
 A:Title: Primary structure and developmental expression of Fbn-1, the mouse fibrillin  
 A:Reference number: A53624; MUID:95130561  
 A:Accession: A53624  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2871 <YIN>  
 A:Cross-references: GB:L29454; NID:g575509; PIDN:AA56840.1; PID:g575510  
 C:Genetics:  
 A:Gene: Fbn-1  
 C:Superfamily: unassigned EGF-related proteins; EGF homology  
 F:1201-1236/Domain: EGF homology <EGF>

Query Match 9.1%; Score 88.5; DB 2; Length 2871;  
 Best Local Similarity 20.9%; Pred. No. 8.8;  
 Matches 42; Conservative 20; Mismatches 70; Indels 69; Gaps 9;

OY 1 MAPHPSGLTTLVPMALALLALGERALALPEICT--QC---PGSVONLSKVA----- 49  
 Db 2389 LCPHGRGFMN-----GADIDCKVIHDVCRNGECVNDKRSYHCICKTGTPDIT 2438  
 OY 50 -----FYCKTTRELMHARCLNOKGTIIGLDLQNC-SLEDGPNFH 90  
 Db 2439 GTSCVDLNECNOAPKPCNFICKNTE---GSYOCSPKGYIIQEDGRSKDLDECATKQH 2494  
 OY 91 QAHFTVILDLQANPLKGLDANLFRGFTQLQTLILPQHNCGGINAMNTITSYDNOICQ 150  
 Db 2495 NCQFLCV-----NTIGGFT-----CKCPGFTQHH--TACIDNNECT 2529  
 OY 151 GOKNLCNNTGDEPMCPENGSC 171  
 Db 2530 SDINLCGSKGICQNTPGSFTC 2550

RESULT 7  
 T04587  
 hypothetical protein F23E13.70 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Sep-1999  
 C:Accession: T04587  
 R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse,  
 submitted to the Protein Sequence Database, March 1998  
 A:Reference number: T15378  
 A:Accession: T04587  
 A:Molecule type: DNA  
 A:Residues: 1-1134 <BEV>  
 A:Cross-references: EMBL:AL022141  
 A:Experimental source: cultivar Columbia; BAC clone F23E13  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 960/3  
 A:Note: F23E13.70  
 C:Superfamily: protein kinase xaz1; leucine-rich alpha-2-glycoprotein repeat homology

Query Match 8.9%; Score 86.5; DB 2; Length 1134;  
 Best Local Similarity 24.6%; Pred. No. 4.9;

```

Matches      50,  Conservative      27,  Mismatches      91,  Indels      35,  Gaps      8;

QY      1  MAPHGPGSLITL-----VFWAAILALALGYERALALPEICTQCPGVSQNLK  47
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      537  VVPGFSSVLVRLRYVNLNSNSFSGEIPOTFGFLRL-VLSLSDHNIIGSIIPETIGNSA  595
QY      48  VAFCKTRELMLNARCOLNKGITLGLD--QNCSLDPGPNFQAHVTYVLIIDQANPL  105
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      596  LEVELRNRLMGHPADLSRLPRLKVIDLDGONNLGELP-PEISQSSLSNLSLDHNL  654
QY      106  KGDLANFERGFTQLOTLLIPQVNCPGGIA-----WNTITSYIDNQIOGQKN  154
      :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      655  SGVIGPSFSGSLNLTKMLSVN-NLTGEIPASIALISSNLVYFNSSNNLKGEIPASIGS  713
QY      155  LCNNT----GDPEMC--PENGSC  171
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      714  RINNTSEFSGNTIELCGKPLNRRC  736

RESULT      8
A55567
Fibrillin I - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 11-Jan-2000
C:Accession: A55567
R:Ritter, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.
Genomics 23, 480-485, 1994
A:Title: Sequence of the coding region of the bovine fibrillin cDNA and localization to
A:Reference number: A55567; MUID:95137597
A:Accession: A55567
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-2871 <TIL>
A:Cross-references: GB:I28748; NID:9508427; PIDN:AAA74122.1; PID:9508428
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:1201-1236/Domain: EGF homology <EGF>

```

```

Query Match Similarity      8.9%  Score 86.5:  DB 2,  length 2871;
Best Local Similarity      20.4%:  Pred. No. 14;
Matches 41:  Conservative 21;  Mismatches 70;  Indels 69;  Gaps

```

---

```

QY      1  NAPHGPGSLTTLVPMAALLLALGVERALALPEICT--QC---PGSVQNLKVA----- 49
      :  |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      2389  LCPHGRGEFWTN-----GADIDECKVIDVDCNMGECVNDRGSYHCITCTGYTPDIT 2438
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      50  -----FYCKTTRELMHARCCLEKQKSTIIGLDPLQNC-STEDPGRNPH 90
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      2439  GTACVDLNECNOAPKPCNTKANTE----GSYQCSCKPKYIIILQEDRSRCKDDECATKQH 2494
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      91  QAHTTVIIDLQANPLKGLDANTFRGTLQTLTLPLQHVNCPGGINAMNTITSYIDNQLQ 150
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      2495  NCOFLCV-----NTIGSFT-----CKCPGFTQHN--TACIDNNECT 2529
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      151  GOKNLNNTGDEPMCEMGESC 171
      |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      2530  SDINLGSKGICQNTPGSFTC 2550

```

---

```

RESULT      9
T23167
hypothetical protein K01C8.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23167
R:Simms, M.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z19702
A:Accession: T23167
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-389 <M1>
A:Cross-References: EMBL:Z49068; PIDN:CA88855.1; GSPDB:GN00020; CESP:K01C8.2

```

A:Experimental source: clone K01C8  
C:Genetics:  
A:Gene: CESP:K01C8.2  
A:Map position: 2  
A:Introns: 54/2; 146/3; 208/3; 283/1; 379/3

Query Match	8.88;	Score 85.5;	DB 2;	Length 389;
Best Local Similarity	25.38;	Pred. No. 1.9;		
Matches 41;	Conservative 25;	Mismatches 65;	Indels 31;	Gaps 11.

OY 81 SLEPPGPNF-----CAHTTVLLID-LOANPLKGLDANTFRGTOLQTL----LPDHVNC 131  
Db 188 IALITNGAL---EMCT-TPGT--OCCSAGYTCQLSVLTAIVYCCGGSTGNTG-----C 236E  
OY 21 LAGVERALALEITOCPSGVOMLSKVAAYCKRYTFLMLHMAKCLLNAGA ILGSDQNC 60  
Db 237 A--DGRVVYQQLAGGYTCETISATSCSPSGDCAPASDPEDFYDVCCITGSTPIENLSCP 294A  
OY 132 GGINAMTTTSYLNDNIOCGOGKNLCNNNGDEMKPENGSGVCP 173  
Db 295 TG---WMSYKNEYDNAN-----RRTCTAVLDIYS-CPIGISCAP 327

RESULT 10  
H84455  
probable receptor-like protein kinase [imported] - Arabidopsis thaliana  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C.Accession: H84455  
R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vannken, S.E.; Umayam, L.; Tallon,  
euss, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A.Reference number: A84420; MUID:20083487  
A.Accession: H84455  
A.Status: preliminary  
A.Molecule type: DNA  
A.Residues: 1-851 <STD>  
A.Cross-references: GB:AE002093; NID:g4689473; FID:AAD27909.1; GSPDB:GN00139  
C.Genetics:  
A.Gene: At2g04300  
A.Map position: 2

	Query Match	8.5%	Score 82.5:	DB 2:	Length 851:	
	Best Local Similarity	23.5%:	Ped. No. 8.5:			
	Matches 42:	Conservative	32:	Mismatches 68:	Indels 37:	Gaps 9
Oy	23 LALERAIAL-----PEICTOCPG--SVONLSKVAFFCKTRELMLHARCINOKGTLIG	74	:       :	:   :   :	:	:
Db	307 LALERLEEFYIDPFELTNODDYIAIKNIQTVSVSTSMQ---- <td>362</td> <td>  :   :   :</td> <td>  :   :   :</td> <td>:</td> <td>:</td>	362	:   :   :	:   :   :	:	:
Oy	75 LDLONCSEDDPGBNQHQAHTVIIDLQANPLKGDIANTRGFQLOLTIIPQHVNCPSGI	134	:   :   :	:   :   :	:	:
Db	363 LNCNNYSIISTP-----PTITFLNLSSSHLGIIASATONLIHLIONDLNSN-NLTGV	414	:   :   :	:   :   :	:	:
Oy	135 NAM-----NNITSYINDQIOGGKNLCNNTGDPDM-CPENGSCV-PPCGP	177	:   :   :	:   :   :	:	:
Db	415 PELLAGLKSLVLINSGNNISGSPOTLLQKKKLKLNLEGNIYLNCPD-GSCYSKONG	472	:   :   :	:   :   :	:	:

RESULT 11  
165967  
disintegrin-like metalloproteinase (EC 3.4.24.-), splice form 2 - human  
N:Alternate names: ADAM11; MDC  
C:Species: Homo sapiens (man)  
C:Date: 29-MAY-1998 #sequence\_revision 17-Mar-2000 #text\_change 26-May-2000  
C:Accession: I65967.38539  
R:Katagiri, T.; Harada, Y.; Emi, M.; Nakamura, Y.  
Cytogenet. Cell Genet. 68, 39-44, 1995  
A:Title: Human metalloprotease/disintegrin-like (MDC) gene: exon-intron organization

A:Reference number: 152965; MUID:95044425  
 A:Accession: 165967  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 7-670 <KAT>  
 A:Cross-references: GB:D31872; NID:9505090; PIDN:BA06670.1; PID:01007242; PID:983683  
 R:Emil, M.; Katagiri, T.; Harada, Y.; Saito, H.; Inazawa, J.; Ito, I.; Kasumi, F.; Nakamu  
 Nature Genet. 5, 151-157, 1993  
 A:Title: A novel metalloprotease/disintegrin-like gene at 17q21.3 is somatically rearran  
 A:Reference number: S38539; MUID:94073190  
 A:Accession: S38539  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-225, 'N', 227-495, 'POGRAWLPLCQHLMSSSARPGRHQ' <EMI>  
 A:Cross-references: GB:D17390; NID:9452188; PIDN:BA04213.1; PID:01004732; PID:9484255  
 C:Comment: For an alternative splice form, see PIR:S38539.  
 C:Genetics:  
 A:Gene: GDB:MDC: ADAM11  
 A:Cross-references: GDB:230267; OMIM:155120  
 A:Map position: 17q21.3-17q21.3  
 A:Introns: 28/3; 57/2; 82/3; 105/1; 127/3; 152/3; 176/3; 232/2; 260/3; 291/1; 308/2; 341  
 C:Superfamily: disintegrin homology  
 C:Keywords: alternative splicing; hydrolase; metalloproteinase  
 F:344-427/Domain: disintegrin homology <DIS>

Query Match 8.4%; Score 82; DB 2; Length 670;  
 Best Local Similarity 22.6%; Pred. No. 7.3;  
 Matches 47; Conservative 25; Mismatches 62; Indels 74; Gaps 14;

QY 31 LPEICT---QCPGVONSKAFYCK-----TTRF-----LMLHA--RCC-- 65  
 Db 413 IAECTGDSGCPNHLKLD--GYCCDHEGRCYGGRCRTRDRCQVLMGHAADRFCE 470  
 QY 66 -LNOKGT-----ILGLDQNSLEDPPGPHQATTVIIDQANPLKGDLANTRGFTQ 118  
 Db 471 KINVESTGSGRGKSGWVQCSKODVLCG-----LLCVNLSGAPRLGDLVDISSVTF 525  
 QY 119 LDTLLPOHVNCPG--INAMNTITSYID-----NOIC-----QG 151  
 Db 526 YHQ---GKELDCRGHVLADSDLSYVEDGTACGPNMLCDHRLCPASAFNFTSCPSG 582  
 QY 152 OKNLNNTGDPMPGPNCSV--PPGPG 177  
 Db 583 ERRICSHG--VCSNEGKICQPDWTG 607

RESULT 12  
 T05335

hypothetical protein F1C12.190 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Sep-1999  
 C:Accession: T05335  
 R:Bevan, M.; Terry, N.; Ardiles, W.; Buysmaert, C.; Dasseville, R.; De Clerck, R.; De  
 ewes, H.W.; Mayer, K.F.X.; Scheller, C  
 submitted to the Protein Sequence Database, April 1998  
 A:Reference number: 215408  
 A:Accession: T05335  
 A:Molecule type: DNA  
 A:Residues: 1-992 <BEV>  
 A:Cross-references: EMBL:AL022224  
 A:Experimental source: cultivar Columbia; BAC clone F1C12  
 C:Genetics:  
 A:Map position: 4  
 A:Introns: 884/1  
 A:Note: F1C12.190  
 C:Superfamily: protein kinase Xa21; leucine-rich alpha-2-glycoprotein repeat homology; F

Query Match 8.4%; Score 82; DB 2; Length 992;  
 Best Local Similarity 20.4%; Pred. No. 11;  
 Matches 40; Conservative 28; Mismatches 52; Indels 76; Gaps 10;

QY 6 PGSLITLVMAALLLALGVERALLALPEICTQCPGVONLSKVAFYCKTTRELMHARCC 65  
 Db 483 PGSLITLVMAALLLALGVERALLALPEICTQCPGVONLSKVAFYCKTTRELMHARCC 65  
 QY 66 LNOKGTILGLD--ONCSLEDPPGPHQATTVIIDQANPLKGDLANTRGFTQ 123  
 Db 515 -----SLKIDSKRNKFSKFP--PEFGDCMSLTVIDLSHNOISQIP-----VOISQIR 562  
 QY 124 LPQHVNCPPGINAMNTITSYIDNOI-----COGOKNLNNT-- 159  
 Db 563 ILNLYNV-----SWNSFQSLPNELGVMKSLTSADFSHNNFSGSVPTSGQSYNNISFL 617  
 QY 160 GDPENC-----PENGS 170  
 Db 618 GNPFLCGFSSNPNCS 633

RESULT 13  
 T05225

extensin homolog F1715.160 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999  
 C:Accession: T05225  
 R:Bevan, M.; Vitale, D.; Liguori, R.; Argiriou, A.; De Simone, V.; Hobeisel, J.; Mewe  
 submitted to the Protein Sequence Database, July 1998  
 A:Reference number: 215404  
 A:Accession: T05225  
 A:Molecule type: DNA  
 A:Residues: 1-699 <BEV>  
 A:Cross-references: EMBL:AL031032  
 A:Experimental source: cultivar Columbia; BAC clone F1715  
 C:Genetics:  
 A:Map position: 4  
 A:Note: F1715.160

Query Match 8.4%; Score 81.5; DB 2; Length 699;  
 Best Local Similarity 21.9%; Pred. No. 8.5;  
 Matches 30; Conservative 20; Mismatches 48; Indels 39; Gaps 5;

QY 40 GSVONLSKVAFYCKTTRELMHARCCLNOKGTILGLDQNSLEDPPGPHQATTVIID 99  
 Db 273 GNMKNLEIVF-----MDNDLGCFPSGK-----KLSNVTVD 306  
 QY 100 LQANPLKGDLANTRGFTQLOTLLPOHVNCPGINAMNTITSYIDNOICQGGKNCNT 159  
 Db 307 ASKNSFTVRLPFTSFGTLVSVEIDI-----SGNKLTLGLVPHNICO--LPNLVNL 354  
 QY 160 GDPENC-----WGSQVDPG 175  
 Db 355 YSYNYFSGSGSCVPPG 371

RESULT 14  
 T45718

receptor-kinase like protein - Arabidopsis thaliana  
 N:Alternate names: protein F1P2.130  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Mar-2000  
 C:Accession: T45718  
 R:Choiene, N.; Roberts, C.; Brotlier, P.; Winkler, P.; Cattolico, L.; Artiguenave, F.;  
 submitted to the Protein Sequence Database, November 1999  
 A:Reference number: 223010  
 A:Accession: T45718  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1011 <CHO>  
 A:Cross-references: EMBL:AL132955  
 A:Experimental source: cultivar Columbia; BAC clone F1P2  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 892/1  
 A:Note: F1P2.130



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:17:27 ; Search time 35.36 Seconds

(without alignments)  
381.304 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_1\_177

Perfect score: 177

Sequence: 1 MAPHGPGSLFTLVPMNAALD.....NTGDPMPGPGSCVDPGPG 177

Scoring table: OLIGO

Searched: 219241 seqs, 76174552 residues

Word size: 0

Total number of hits satisfying chosen parameters: 6208

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database:

1: PIR\_68:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5	2.8	20	2	A60822
2	5	2.8	23	2	I55406
3	5	2.8	23	2	B37843
4	5	2.8	24	2	P00677
5	5	2.8	24	2	S10681
6	5	2.8	24	2	S10682
7	5	2.8	25	2	PC1221
8	5	2.8	25	2	S10850
9	5	2.8	26	2	A32806
10	5	2.8	26	2	B44107
11	5	2.8	26	2	A61056
12	5	2.8	26	2	S05414
13	5	2.8	27	2	I46566
14	5	2.8	28	2	S29135
15	5	2.8	28	2	S29136
16	5	2.8	28	2	A46690
17	5	2.8	30	2	S21814
18	5	2.8	30	2	E45095
19	5	2.8	30	4	S12902
20	4	2.3	10	2	A49581
21	4	2.3	10	2	I48778
22	4	2.3	11	2	A33917
23	4	2.3	11	2	F33098
24	4	2.3	12	2	S25485
25	4	2.3	12	2	C58502
26	4	2.3	12	2	A26093
27	4	2.3	13	2	A38929
28	4	2.3	13	2	A60336
29	4	2.3	13	2	S33800

30	4	2.3	14	2	S33801	chaperone, TCPI-re
31	4	2.3	14	2	B61235	fibroblast-activat
32	4	2.3	14	2	I49514	B144 protein A - m
33	4	2.3	14	2	PH1625	Ig H chain V-D-J r
34	4	2.3	14	2	PH1627	Ig H chain V-D-J r
35	4	2.3	15	2	C37765	hypothetical prote
36	4	2.3	15	2	A08416	lombicrine kinase
37	4	2.3	15	2	A36315	recycling receptor
38	4	2.3	15	2	I56046	urinary tract ston
39	4	2.3	15	2	S29175	D-galactose-bindin
40	4	2.3	16	2	I57530	gene c-fms protein
41	4	2.3	16	2	CORT	cerebellin - rat
42	4	2.3	16	2	B25979	cerebellin - rat
43	4	2.3	16	2	I78533	nodulation protein
44	4	2.3	16	2	PL0124	gene agouti protei
45	4	2.3	16	2	E58501	cerebellin - pig
						25K kidney and gal

#### ALIGNMENTS

RESULT 1  
A60822  
cytochrome P450 PB-3a - rat (fragment)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 30-Sep-1993 #sequence-revision 30-Sep-1993 #text-change 11-Jun-1999  
C/Accession: A60822; I55191  
R:Ameliazad, Z.; Narbonne, J.F.; Wolf, C.R.; Robertson, L.W.; Oesch, F.  
Biochem. Pharmacol. 37, 3245-3249, 1988  
A>Title: Effect of nutritional imbalances on cytochrome P-450 isozymes in rat liver.  
A:Reference number: A60822; MUID:88293549  
A/Accession: A60822  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-20 <AMP>  
R:Heshimoto, T.; Matsmoto, T.; Nishizawa, M.; Kawabata, S.  
J. Biochem. 103, 487-492, 1988  
A>Title: A mutant rat strain deficient in induction of a phenobarbital-inducible for  
A:Reference number: I55191; MUID:8823074  
A/Accession: I55191  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-? <RES>  
A:Cross-references: GB:D00250; MUID:q220727; PIDN:BA00181.1; PID:q220728  
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
C:Keywords: electron transfer; heme; monooxygenase; oxidoreductase; transmembrane pro

Query Match 2.8; Score 5; DB 2; Length 20;  
Best Local Sim.arity 100.0%; Pred. No. 3.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LLLAL 23  
Db 7 LLLAL 11

RESULT 2  
I55406  
nicotinic acetylcholine receptor - Rattus leucopus (fragment)  
C:Species: Rattus leucopus  
C>Date: 02-Jul-1996 #sequence-revision 02-Jul-1996 #text-change 20-Aug-1999  
C/Accession: I55406  
R:Walke, W.; Staple, J.; Adams, L.; Gnegy, M.; Chahine, K.; Goldman, D.  
J. Biol. Chem. 269, 19447-19456, 1994  
A>Title: Calcium-dependent regulation of rat and chick muscle nicotinic acetylcholine  
A:Reference number: I55406; MUID:94308229  
A/Accession: I55406  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-23 <RES>  
A:Cross-references: GB:L10594; MUID:g310180; PIDN:AA21730.1; PID:g310181

C:Genetics:  
 A:Gene: nAChR  
 C:Superfamily: acetylcholine receptor  
 C:Keywords: neurotransmitter receptor

Query Match 2.8%; Score 5; DB 2; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 LLLAL 23  
 |||||  
 Db 10 LLLAL 14

RESULT 3  
 B37843  
 vrg 18 protein - Bordetella pertussis (fragment)  
 C:Species: Bordetella pertussis  
 C:Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 17-Feb-1994  
 C:Accession: B37843  
 R:Beatlie, D.T.; Knapp, S.; Mekalanos, J.J.  
 J. Bacteriol. 172, 6997-7004, 1990  
 A:Title: Evidence that modulation requires sequences downstream of the promoters of two  
 A:Reference number: A37843; MUID:91072251  
 A:Accession: B37843  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-23 <BEA>  
 A:Cross-references: GB:M37228

Query Match 2.8%; Score 5; DB 2; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 AALLL 20  
 |||||  
 Db 10 AALLL 14

RESULT 4  
 P00677  
 photosystem I 9.0K H2 chain - common tobacco (fragment)  
 C:Species: Nicotiana tabacum (common tobacco)  
 C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 17-Mar-1999  
 C:Accession: P00677  
 R:Obokata, J.; Mikami, K.; Hayashida, N.; Nakamura, M.; Sugitara, M.  
 Plant Physiol. 102, 1259-1267, 1993  
 A:Title: Molecular heterogeneity of photosystem I. psad, psae, psaf, psah and psal are  
 A:Reference number: P00667; MUID:94105345  
 A:Accession: P00677  
 A:Molecule type: protein  
 A:Residues: 1-24 <OBO>  
 C:Superfamily: photosystem I protein psah  
 C:Keywords: chloroplast; photosynthesis; photosystem I; thylakoid

Query Match 2.8%; Score 5; DB 2; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 DLANT 112  
 |||||  
 Db 13 DLANT 17

RESULT 5  
 S10681  
 probable 7-ethoxycoumarin O-deethylase (EC 1.14.14.-) cytochrome P450 isoform 3 - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 05-Mar-1999  
 C:Accession: S10681

R:Sinclair, J.F.; Wood, S.; Lambrecht, L.; Gorman, N.; Mende-Mueller, L.; Smith, L.;  
 Biochem. J. 269, 85-91, 1990  
 A:Title: Isolation of four forms of acetone-induced cytochrome P-450 in chicken liver  
 A:Reference number: S10680; MUID:90328998  
 A:Accession: S10681  
 A:Molecule type: protein

A:Residues: 1-24 <SIN>  
 C:Superfamily: unassigned cytochrome P450; cytochrome P450 homology  
 C:Keywords: electron transfer; endoplasmic reticulum; heme; membrane protein; monooxy

Query Match 2.8%; Score 5; DB 2; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AALLL 21  
 |||||  
 Db 5 AALLL 9

RESULT 6  
 S10682  
 cytochrome P450 isoform 4 - chicken (fragment)  
 N:Contains: oxidoreductase (EC 1.-.-.-)  
 C:Species: Gallus gallus (chicken)  
 C:Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 05-Mar-1999  
 C:Accession: S10682  
 R:Sinclair, J.F.; Wood, S.; Lambrecht, L.; Gorman, N.; Mende-Mueller, L.; Smith, L.;  
 Biochem. J. 269, 85-91, 1990  
 A:Title: Isolation of four forms of acetone-induced cytochrome P-450 in chicken liver  
 A:Reference number: S10680; MUID:90328998  
 A:Accession: S10682  
 A:Molecule type: protein  
 A:Residues: 1-24 <SIN>  
 C:Superfamily: unassigned cytochrome P450; cytochrome P450 homology  
 C:Keywords: electron transfer; endoplasmic reticulum; heme; membrane protein; monooxy

Query Match 2.8%; Score 5; DB 2; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AALLL 21  
 |||||  
 Db 5 AALLL 9

RESULT 7  
 PC1221  
 alanine racemase (EC 5.1.1.1) - Pseudomonas fluorescens (fragment)  
 C:Species: Pseudomonas fluorescens  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
 C:Accession: PC1221  
 R:Yokoiawa, K.; Kawai, H.; Endo, K.; Lim, Y.H.; Esaki, N.; Soda, K.  
 Biosci. Biotechnol. Biochem. 57, 93-97, 1993  
 A:Title: The molecular structure of alanine racemase from a psychrotroph, Pseudomonas fluorescens:  
 A:Reference number: PC1221; MUID:93169026  
 A:Accession: PC1221  
 A:Molecule type: protein  
 A:Residues: 1-25 <YOK>  
 C:Comment: This enzyme catalyzes the racemization of L- and D-alanine.  
 C:Superfamily: alanine racemase  
 C:Keywords: isomerase; pyridoxal phosphate

Query Match 2.8%; Score 5; DB 2; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 IDLOA 102  
 |||||  
 Db 8 IDLOA 12

## RESULT 8

S10850

alpha-amylase inhibitor - durum wheat (fragment)

N:Alternate names: glutenin low molecular weight chain

C:Species: Triticum durum (durum wheat)

C&gt;Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997

C:Accession: S10850

R:Kobrehel, K.; Alary, R.

J. Sci. Food Agric. 48, 441-452, 1989

A:Title: Isolation and partial characterisation of two low molecular weight durum wheat

A:Reference number: S10849

A:Accession: S10850

A:Molecule type: protein

A:Residues: 1-25 &lt;KOB&gt;

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: alpha-amylase inhibitor

## Query Match

Best Local Similarity 100.0%; Score 5; DB 2; Length 25;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GSCVP 173

DB 2 GSCVP 6

## RESULT 9

transposase - Escherichia coli transposon Tn4651 (fragment)

C:Species: Escherichia coli

C&gt;Date: 20-Oct-1989 #sequence\_revision 20-Oct-1989 #text\_change 20-Jun-2000

C:Accession: A32806

R:Tsuda, M.; Minegishi, K.I.; Iino, T.

J. Bacteriol. 171, 1386-1393, 1989

A:Title: Toluene transposons Tn4651 and Tn4653 are class II transposons.

A:Reference number: A32806; MUID:89155438

A:Accession: A32806

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-26 &lt;TSU&gt;

A:Cross-references: GB:D90148; MID:9217162; PIDN:BA14178.1; PID:9217164

C:Superfamily: transposase Tn21

## Query Match

Best Local Similarity 100.0%; Score 5; DB 2; Length 26;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 LALPE 33

DB 16 LALPE 20

## RESULT 10

B44107

cytochrome P450 NF3, beta-naphthoflavone-induced (N-terminal) - chicken (fragment)

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Gallus gallus (chicken)

C&gt;Date: 27-Apr-1993 #sequence\_revision 20-Aug-1994 #text\_change 05-Mar-1999

C:Accession: B44107

R:Nakai, K.; Ward, A.M.; Gannon, M.; Rifkind, A.B.

J. Biol. Chem. 267, 19503-19512, 1992

A:Title: Beta-naphthoflavone induction of a cytochrome P-450 arachidonic acid epoxigenase

A:Reference number: A44107; MUID:92406903

A:Accession: B44107

A:Molecule type: protein

A:Residues: 1-26 &lt;NAK&gt;

A:Experimental source: embryo liver

A&gt;Note: sequence extracted from NCBI backbone (NCBIP:113919)

C:Superfamily: unassigned cytochrome P450; cytochrome P450 homology

C:Keywords: electron transfer; heme; monooxygenase; oxidoreductase

## Query Match

Best Local Similarity 100.0%; Score 5; DB 2; Length 26;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 AALL 21

DB 5 AALL 9

## RESULT 11

A61056

aminopyrine N-demethylase (EC 1.14.14.-) cytochrome P450 PB-A - chicken (fragment)

N:Alternate names: 7-ethoxycoumarin O-deethylase; phenobarbital-induced 52 kDa cytoch

C:Species: Gallus gallus (chicken)

C&gt;Date: 31-Dec-1993 #sequence\_revision 03-Feb-1994 #text\_change 11-May-2000

C:Accession: C44107; A61056; S13263

R:Nakai, K.; Ward, A.M.; Gannon, M.; Rifkind, A.B.

J. Biol. Chem. 267, 19503-19512, 1992

A:Title: Beta-naphthoflavone induction of a cytochrome P-450 arachidonic acid epoxyg

A:Reference number: A44107; MUID:92406903

A:Accession: C44107

A:Molecule type: protein

A:Residues: 1-26 &lt;NAK&gt;

A:Experimental source: embryo liver

A&gt;Note: sequence extracted from NCBI backbone (NCBIP:113920)

R:Gupta, R.P.; Lapadula, D.M.; Abou-Donia, M.B.

Comp. Biochem. Physiol. C 96, 163-176, 1990

A:Title: Purification and characterization of cytochrome P-450 isozymes from phenobar

A:Reference number: A61056; MUID:91130218

A:Accession: A61056

A:Molecule type: protein

A:Residues: 1-19 &lt;GUP&gt;

R:Gupta, R.P.; Lapadula, D.M.; Abou-Donia, M.B.

Arch. Biochem. Biophys. 282, 170-182, 1990

A:Title: Purification and characterization of cytochrome P450 isozymes from beta-naph

A:Accession: S13263; MUID:91024193

A:Molecule type: protein

A:Residues: 1-24 &lt;GUP&gt;

C:Superfamily: unassigned cytochrome P450; cytochrome P450 homology

C:Keywords: electron transfer; endoplasmic reticulum; heme; membrane protein; monooxy

## Query Match

Best Local S. 100.0%; Score 5; DB 2; Length 26;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 AALL 21

DB 5 AALL 9

## RESULT 12

S05414

steryl-sulfatase (EC 3.1.6.2), microsomal - rat (fragment)

N:Alternate names: arylsulfatase C

C:Species: Rattus norvegicus (Norway rat)

C&gt;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 28-Apr-1993

C:Accession: S05414

R:Kawanu, J.I.; Kotani, T.; Ohtaki, S.; Minamino, N.; Matsuo, H.; Oinuma, T.; Aikawa,

Biochim. Biophys. Acta 997, 199-205, 1989

A:Title: Characterization of rat and human steroid sulfatases.

A:Reference number: S05414; MUID:89352671

A:Accession: S05414

A:Molecule type: protein

A:Residues: 1-26 &lt;KAW&gt;

C:Keywords: sulfuric ester hydrolase

Query Match 2.8%; Score 5; DB 2; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 85 GPNP 89  
 |||||  
 Db 5 GPNP 9

RESULT 13  
 I46566  
 (antileukoprotease - pig (fragment))  
 C:Species: Sus scrofa domestica (domestic pig)  
 C>Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 21-Jul-2000  
 C:Accession: I46566  
 R:Stimmen, R.C.; Badinga, L.; Michel, F.J.  
 Mol. Cell. Endocrinol. 97, 101-108, 1993  
 A:Title: Chromosomal organization of the gene encoding porcine antileukoprotease and A  
 A:Reference number: I46566; MUID:94192847  
 A:Accession: I46566  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-27 <SIM>  
 A:Cross-references: GB:I20836; NID:g1182068; PIDN:AA86935.1; PID:g520360  
 C:Superfamily: antileukoprotease; antileukoprotease repeat homology

Query Match 2.8%; Score 5; DB 2; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LLALG 24  
 |||||  
 Db 11 LLALG 15

RESULT 14  
 S29135  
 aminopyrine N-demethylase (EC 1.14.14.-) cytochrome P450 betanF-A1 - chicken (fragment)  
 N:Alternate names: 7-ethoxycoumarin O-deethylase  
 C:Species: Gallus gallus (chicken)  
 C>Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 05-Mar-1999  
 C:Accession: S29135  
 R:Gupta, R.P.; Lapadula, D.M.; Abou-Donia, M.B.  
 Arch. Biochem. Biophys. 282, 170-182, 1990  
 A:Title: Purification and characterization of cytochrome P450 isozymes from beta-naphtho  
 A:Reference number: S13263; MUID:91024193  
 A:Accession: S29135  
 A:Molecule type: protein  
 A:Residues: 1-28 <GUP>  
 C:Superfamily: unassigned cytochrome P450; cytochrome P450 homology  
 C:Keywords: electron transfer; endoplasmic reticulum; heme; membrane protein; monooxygen

Query Match 2.8%; Score 5; DB 2; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 AALLL 21  
 |||||  
 Db 5 AALLL 9

RESULT 15  
 S29136  
 aminopyrine N-demethylase (EC 1.14.14.-) cytochrome P450 betanF-A2 - chicken (fragment)  
 N:Alternate names: 7-ethoxycoumarin O-deethylase  
 C:Species: Gallus gallus (chicken)  
 C>Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 05-Mar-1999  
 C:Accession: S29136  
 R:Gupta, R.P.; Lapadula, D.M.; Abou-Donia, M.B.  
 Arch. Biochem. Biophys. 282, 170-182, 1990  
 A:Title: Purification and characterization of cytochrome P450 isozymes from beta-naphtho

A:Reference number: S13263; MUID:91024193  
 A:Accession: S29136  
 A:Molecule type: protein  
 A:Residues: 1-28 <GUP>  
 C:Superfamily: unassigned cytochrome P450; cytochrome P450 homology  
 C:Keywords: electron transfer; endoplasmic reticulum; heme; membrane protein; monooxy

Query Match 2.8%; Score 5; DB 2; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

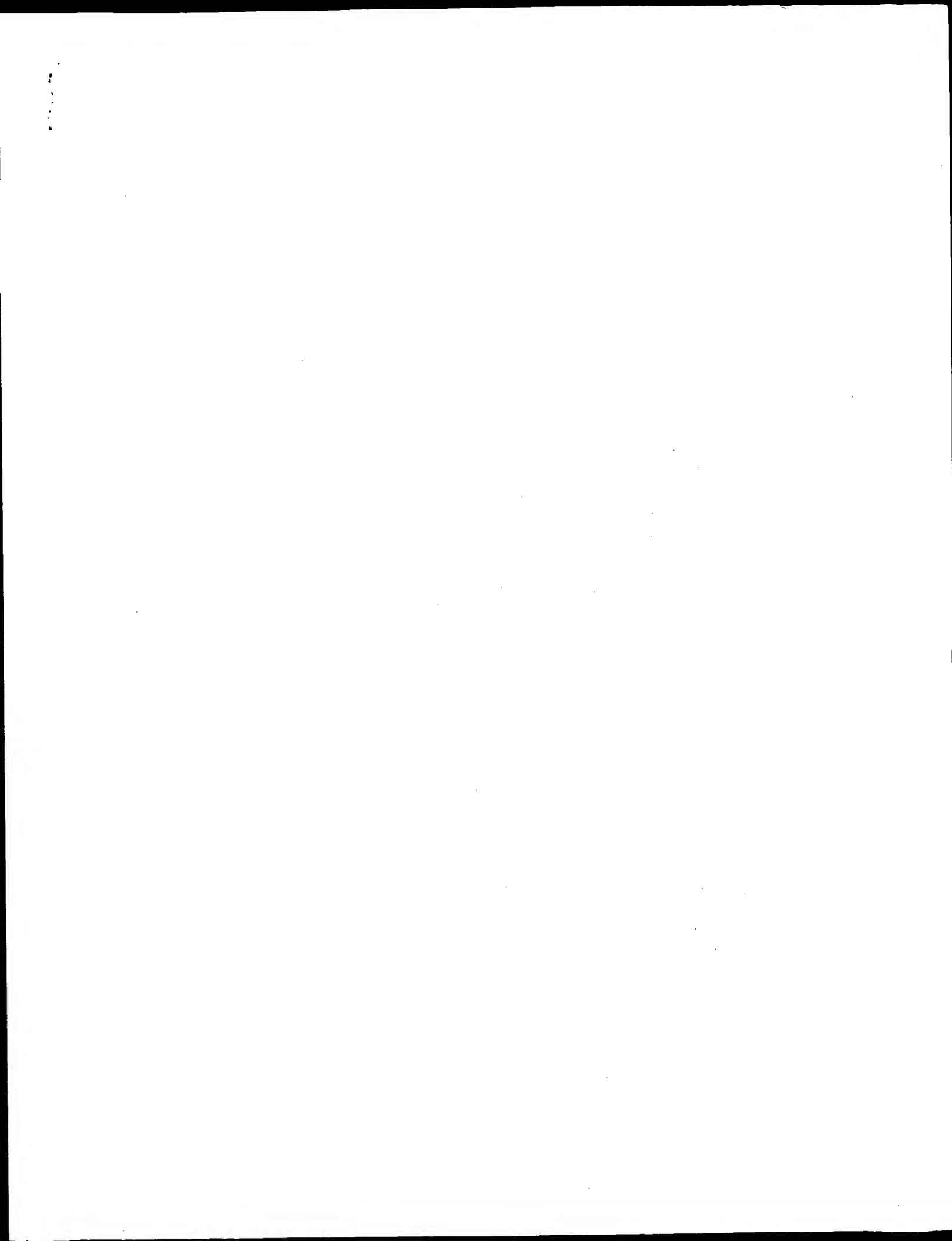
QY 17 AALLL 21  
 |||||  
 Db 5 AALLL 9

Search completed: March 28, 2002, 09:17:28  
 Job time: 382 sec



Thu Mar 28 09:21:41 2002

us-09-726-348-2\_copy\_1\_177.rpx



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:01:34 ; Search time 74.25 seconds  
(without alignments)  
53.348 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_126\_177  
Perfect score: 309  
Sequence: 1 QHVNCPGGINAMNTITSYD.....NTGDEPMCPENSCVPDPG 52

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR-68:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	67	21.7	2019	2 A33996	sodium channel pro
2	66.5	21.5	1820	1 CHEE	sodium channel pro
3	66	21.4	1599	2 T16210	hypothetical prote
4	64.5	20.9	1101	2 T16840	hypothetical prote
5	63.5	20.6	886	2 A57172	probable hormone r
6	63	20.4	2476	2 T34022	zonadhesin - pig
7	63	20.4	3002	2 A47221	fibrillin 1 precur
8	62.5	20.2	1964	2 T09059	notch4 - mouse
9	61.5	19.9	746	2 S27985	cellulase (EC 3.2.
10	61.5	19.9	2005	2 B25019	sodium channel pro
11	61.5	19.9	2907	2 A57278	fibrillin-2 precur
12	60.5	19.6	1835	2 I54323	sodium channel alp
13	60.5	19.6	1836	2 JS0648	sodium channel alp
14	60.5	19.6	1965	2 T33215	hypothetical prote
15	60.5	19.6	2005	2 A46269	sodium channel alp
16	60.5	19.6	2871	2 A55367	fibrillin 1 - bovl
17	60	19.4	48	2 S29216	neurotoxin Tx2 - s
18	60	19.4	260	1 VCBW	coat protein - sou
19	60	19.4	279	1 VCBWSC	coat protein - sou
20	60	19.4	1951	2 S00370	sodium channel pro
21	60	19.4	1983	2 A60054	sodium channel pro
22	60	19.4	2871	2 A55624	fibrillin-1 precur
23	59.5	19.3	319	2 G85818	hypothetical prote
24	59.5	19.3	389	2 T23167	hypothetical prote
25	59.5	19.3	495	2 B70322	hypothetical prote
26	59.5	19.3	568	2 H86552	prolyl tRNA synthe
27	59.5	19.3	568	2 F72070	proline--tRNA 119a
28	59.5	19.3	1335	2 H86489	protein T32E20.13
29	59.5	19.3	1353	1 J02168	E2 glycoprotein pr

30	59.5	19.3	1361	2 S29998	surface protein -
31	59.5	19.3	1362	2 A37474	surface glycoprote
32	59.5	19.3	2098	2 T18397	protein CTRP - mal
33	59.5	19.1	49	2 S29215	neurotoxin Tx2 - s
34	59	19.1	779	2 T20654	hypothetical prote
35	59	19.1	1155	2 S64365	GDP/GTP exchange p
36	59	19.1	1522	2 J01101	sodium channel pro
37	59	19.1	1895	2 T15881	hypothetical prote
38	59	19.1	2818	2 A54105	fibrillin-2 precur
39	58.5	18.9	1840	1 CHRTM1	sodium channel pro
40	58.5	18.9	2180	2 T29764	hypothetical prote
41	58.5	18.9	2352	2 T30201	Notch homolog prot
42	58.5	18.9	2643	2 T29149	hypothetical prote
43	58	18.8	677	2 C42125	trophozoite cystei
44	58	18.8	2016	2 A38195	sodium channel pro
45	58	18.8	2664	2 T28626	variant-specific s

## ALIGNMENTS

```

RESULT 1
A33996
Sodium channel protein I, cardiac - rat
N:Alternate names: sodium channel protein (SKM2) alpha chain
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 20-Aug-1999
C:Accession: A33996; J00412
R:Rogart, R.B.; Crittbs, L.L.; Muglia, L.K.; Kephart, D.D.; Kaiser, M.W.
Proc. Natl. Acad. Sci. U.S.A. 86, 8170-8174, 1989
A>Title: Molecular cloning of a putative tetrodotoxin-resistant rat heart Na(+) chann
A:Reference number: A33996; MUID:90046760
A:Accession: A33996
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2019 <K0G>
A:Cross-references: GB:M27902; NID:9206857; PIDN:AAA42114.1; PID:9206858
R:Kallen, R.G.; Sheng, Z.H.; Yang, J.; Chen, L.; Rogart, R.B.; Barchi, R.L.
Neuron 4, 233-242, 1990
A>Title: Primary structure and expression of a sodium channel characteristic of dener
A:Reference number: J00412; MUID:90166613
A:Accession: J00412
A:Molecule type: mRNA
A:Residues: 1-479,481-1712,'T',1714-1963,'R',1965-2019 <KAL>
A:Superfamily: sodium channel protein
C:Experimental source: muscle
C:Keywords: cardiac muscle; duplication; heart; sodium channel; transmembrane protein

Query Match          21.7% Score 67; DB 2; Length 2019;
Best Local Similarity 29.6% Pred.No. 27;
Matches 16: Conservative 8; Mismatches 20; Indels 10; Gaps 2;

QY      7 GGINA---WNTTXYIDNOICOGORN-----LCNNTGDEPMCPENSCVPDPG 50
          | | | | | | | | | | | | | | | | | | | | | | | | |
Db      293 GSEVADGIVWNSLDVYLNDPANYLKNGTDTVLGNSSDAGTCPGYCNLAAG 346

RESULT 2
CHEE
Sodium channel protein - electric eel
C:Species: Electrophorus electricus (electric eel)
C>Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 22-Jun-1999
C:Accession: A03178; I50536
R:Noda, M.; Shimizu, S.; Tanabe, T.; Takai, T.; Kayano, T.; Ikeda, T.; Takahashi, H.;
da, H.; Miyata, T.; Numa, S.
Nature 312, 121-127, 1984
A>Title: Primary structure of Electrophorus electricus sodium channel deduced from cd
A:Reference number: A03178; MUID:85061498
A:Accession: A03178
A:Molecule type: mRNA
A:Residues: 1-1820 <NDP>
A:Cross-references: GB:X01119; NID:962776; PIDN:CAA25587.1; PID:962777

```

R:Noda, M.; Numa, S.

J. Recept. Res. 7, 467-497, 1987

A:Title: Structure and function of sodium channel.

A:Reference number: 150536; MUID:87311393

A:Accession: 150536

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1820 <MO>

A:Cross-references: GB:M2252; NID:g1041048; PIDN:AAA79960.1; PID:g1041049

C:Comment: This membrane glycoprotein mediates the voltage-dependent sodium-ion permeability of the membrane, the protein forms a sodium-selective channel through which sodium ions

C:Comment: This sequence contains four highly homologous internal repeats (excluding repeats 1 and 2) which are thought to be oriented pseudosymmetrically across the membrane.

C:Comment: Each repeat has a similar overall structure containing six subregions located in identical positions relative to the overall structure.

C:Comment: The four repeating units are thought to be oriented pseudosymmetrically across the membrane.

C:Comment: The presence of four homologous structures within this molecule is consistent with the

C:Comment: Available data suggest that activation and inactivation gates are located near

C:Comment: 955 might, in conjunction with the positively charged residues of S4, act as a voltage sensor.

C:Keywords: duplication; glycoprotein; ion transport; membrane protein; sodium channel;

C:Keywords: duplication; glycoprotein; ion transport; membrane protein; sodium channel;

F:111-419,555-807,988-1281,1311-1587/Region: duplication internal repeats I, II, III and

F:111-41,555-585,988-1019,1311-1341/Region: S1 of repeats I through IV

F:150-171,597-620,1053-1057,1353-1376/Region: S2 of repeats I through IV

F:177-197,626-643,1062-1079,1381-1398/Region: S3 of repeats I through IV

F:204-224,651-671,1092-1112,1417-1437/Region: S4 of repeats I through IV

F:244-264,691-711,1132-1152,1454-1474/Region: S5 of repeats I through IV

F:379-402,767-790,1236-1264,1544-1567/Region: S6 of repeats I through IV

F:205,278,288,317,591,690,797,1160,1174,1806/Binding site: carbonyl (Asn) (covalent)

F:205,278,288,317,591,690,797,1160,1174,1806/Binding site: carbonyl (Asn) (covalent)

Query Match 21.5%; Score 66.5; DB 1; Length 1820;

Best Local Similarity 34.0%; Pred. No. 27;

Matches 17; Conservative 8; Mismatches 16; Indels 9; Gaps 3;

QY 10 NAMNT---ITSYDNO---ICGOKN-ICNNTGDEPCNGSCVPDG 50

Db 285 SAYNTFETAYIENEENQYFLDGDALDALLCGNSDAGKCEGYTCMKAG 334

RESULT 3

T16210 hypothetical protein F30H5.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T16210

R:Pauley, A.; Stellyes, L.

submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid F30H5.

A:Reference number: 218478

A:Accession: T16210

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1599 <PAU>

A:Cross-references: EMBL:U29096; NID:g861390; PID:g861393; PIDN:AAA6408.1; CESP:F30H5.3

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F30H5.3

A:introns: 12/1; 59/2; 85/3; 124/3; 217/2; 534/3; 560/1; 1549/1

Query Match 21.4%; Score 66; DB 2; Length 1599;

Best Local Similarity 27.7%; Pred. No. 27;

Matches 18; Conservative 6; Mismatches 25; Indels 16; Gaps 2;

QY 2 HANCPGGINAM-----NTITSYDNOIC-----GOKNLCNNTGDEPCNGS 45

Db 593 YTGCGNDNNFETLECCNTCENIIPPOCGQDAYKDYOGNYVCNSGAGNSCPVNYE 652

QY 46 CVPDG 50

Db 653 CYFDG 657

RESULT 4

T16840

hypothetical protein T10E10.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T16840

R:Geisel, C.

submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid T10E10.

A:Reference number: 218588

A:Accession: T16840

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1101 <GEI>

A:Cross-references: EMBL:U39644; NID:g1049339; PID:g1049343; PIDN:AAA80360.1; CESP:T1

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:T10E10.4

A:introns: 93/2; 152/2; 191/3; 209/2; 283/3; 303/1; 399/3; 421/1; 440/1; 465/1; 547/3

Query Match 20.9%; Score 64.5; DB 2; Length 1101;

Best Local Similarity 26.6%; Pred. No. 28;

Matches 17; Conservative 6; Mismatches 14; Indels 27; Gaps 3;

QY 4 NCPGGINAMNTITSYDNOIC-----GOKNLCNNTGDP-----EMCPENG 44

Db 390 SCPG-----TSCQCNKVCQOQDTLNLNLNIOHAPLCGSNVPPLGSCNEQCFQYS 441

QY 45 SCVP 48

Db 442 ACTP 445

RESULT 5

A57172

probable hormone receptor EMR1 precursor - human

C:Species: Homo sapiens (man)

C:Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 11-Jan-2000

C:Accession: A57172

R:Band, V.; Chissoe, S.L.; Viegas-Pequignot, E.; Diriong, S.; Nguyen, V.C.; Roe, B.A

Genomics 26, 334-344, 1995

A:Title: EMR1, an unusual member in the family of hormone receptors with seven transmembrane domains

A:Reference number: A57172; MUID:95324926

A:Accession: A57172

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-886 <BAU>

A:Cross-references: GB:X81479; NID:g784993; PIDN:CAA57232.1; PID:g784994

C:Genetics:

A:Gene: GDB:EMR1

A:Cross-references: GDB:378349; OMIM:600493

A:Map position: 19p13.3-19p13.3

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: transmembrane protein

F:136-170/Domain: EGF homology <EGF>

Query Match 20.6%; Score 63.5; DB 2; Length 886;

Best Local Similarity 30.0%; Pred. No. 29;

Matches 18; Conservative 7; Mismatches 22; Indels 13; Gaps 3;

QY 5 CPGGINAMNTITSY-----IDNOCGOKNLCNNTGD-PEMCPENGSCVPDPG 52

Db 182 CEPHATCNTVGNISCFNCPGFESSGHLSCGLKASCEIDIECTEMKPINSTCT-NTPG 240

RESULT 6

T34022

zonadhesin - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000

C:Accession: T34022

R.Hardy, D.M.; Garbers, D.L.  
J. Biol. Chem. 270, 26025-26028, 1995

A.Title: A sperm membrane protein that binds in a species-specific manner to the egg extracellular matrix  
A.Reference number: Z1464; MUID:96064658

A.Accession: T34022

A.Status: preliminary; translated from GB/EMBL/DDBJ

A.Molecule type: mRNA

A.Residues: 1-2476 <HAR>

A.Cross-references: EMBL:U40024; NID:g1066465; PID:g1066466; PIDN:AAC48486.1

A.Experimental source: strain Weishan; testis

C.Genetics:

A.Gene: Zan

C.Function:

A.Description: may be involved in sperm adhesion to the zona pellucida

Query Match                      20.4%; Score 63; DB 2; Length 2476;  
Best Local Similarity          30.4%; Pred. No. 87;  
Matches    14; Conservative     6; Mismatches    24; Indels        2; Gaps        1;

OY                      5 CPGGINANNNTTSTYDNOI--CQGQKNLCNNTGDPENCPENGSCVP 48  
Db                      1847 CPLDSAHSVYTCVPCSLPSCODPEGQCAGAPSTCEGCIPEP 1892  
  
11 :  
:: : : : : : : : : : : : : : : : : : : :

RESULT                      7  
AA7221  
fibritin 1 precursor - human (fragment)  
C.Species: Homo sapiens (man)  
C.Date: 02-Jun-1995 #sequence revision 25-Apr-1997 #text change 21-Jul-2000  
C.Accession: AA7221; I54555; S17064; I59574; S17062; S62111; A34198  
R.Corson, G.M.; Chalderg, S.C.; Dietz, H.C.; Charbonneau, N.W.; Sakai, L.Y.  
Genomics 17, 476-484, 1993  
A>Title: Fibritin binds calcium and is coded by cDNAs that reveal a multidomain structural organization  
A.Reference number: AA7221; MUID:94010947  
A.Accession: AA7221  
A:Molecule type: mRNA  
A.Residues: 1-337,'T','T',339-1029 <COR>  
A.Cross-references: GB:X63556  
R.Perella, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Panglilan, T.; Bonad  
Hum. Mol. Genet. 2, 961-968, 1993  
A>Title: Genomic organization of the sequence coding for fibritin, the defective gene f  
A.Reference number: I54355; MUID:93372860  
A.Accession: I54355  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A.Residues: 132-3002 <PER>  
A.Cross-references: GB:I13923; NID:g306745; PIDN:ABO2036.1; PID:g306746  
R.Maslen, C.L.; Corsen, G.M.; Maddox, B.K.; Glaville, R.W.; Sakal, L.Y.  
Nature 352, 334-337, 1991  
A>Title: Partial sequence of a candidate gene for the Marfan syndrome.  
A.Reference number: S17064; MUID:91304568  
A.Accession: S17064  
A:Molecule type: mRNA  
A.Residues: 1030-3002 <MAS>  
A.Cross-references: EMBL:X63556  
R.Dietz, H.C.; Valle, D.; Francomano, C.A.; Kendzior, R.J.  
Science 259, 680-683, 1993  
A>Title: The skipping of constitutive exons in vivo induced by nonsense mutations.  
A.Reference number: I59574; MUID:93157831  
A.Accession: I59574  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A.Residues: 2217-2288,'T',2290-2325 <RES>  
A.Cross-references: GB:S54426; NID:g264860; PIDN:ABE25244.1; PID:g264861  
R.Ilee, B.; Godfrey, M.; Vitale, E.; Horri, H.; Mattei, M.G.; Sartarazi, M.; Tsipouras, P.  
Nature 332, 330-334, 1991  
A>Title: Linkage of Marfan syndrome and a phenotypically related disorder to two differ  
A.Reference number: S17062; MUID:91304567  
A.Accession: S17062  
A:Molecule type: mRNA  
A.Residues: "LVATVFIFILSYNKML",944-1444 <LEE>  
A.Cross-references: EMBL:X62006; NID:g31398; PIDN:CAB56534.1; PID:g5924015

```

A:Accession:S62111
A:Molecule type: protein
A:Residues: 1166-1176,'X',1178-1180,'D',1182-1185 <LEE2>
R:Maddox, B.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W.
J. Biol. Chem. 264, 21381-21385, 1989
A>Title: Connective tissue microfibrils. Isolation and characterization of three large
A:Reference number: A34198; MUID:90078246
A:Accession: A34198
A:Molecule type: protein
A:Residues: 565-575;1890-1892,'I',1894-1900 <MAD>
C:Comment: Fibrillin is a major component of elastin-associated microfibrills.
C:Genetics:
A:Gene: GDB:FBN1
A:Cross-references: GDB:127115; OMTM:134797; OMTM:154700
A:Map position: 15q21.1-15q21.1
A:Introns: 2236/1; 2258/1; 2297/1
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein
F:1-1002/Product: fibrillin (5'-region exon C splice form) #status predicted <MATCH>
F:1332-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MATCH>
F:1332-1367/Domain: EGF homology <EGF>
F:1457-1492/Domain: EGF homology <EGF>
F:2262-2295/Domain: EGF homology <EGF1>

Query Match          20.4%; Score 63; DB 2; Length 3002;
Best Local Similarity 35.7%; Pred. No. 1e+02;
Matches 15; Conservative 5; Mismatches 20; Indels 2; Gaps 1;

QY      5  CPGGINAMNTITSYIDNOICOGKRLCNTGDPKCPENGSC 46
        ||| :||| | | | | | | | | | | | | | | | |
Db       2642 CPPGFTQH--TSCIDNNECTSDINLGSKGICQMPGSPFC 2681

RESULT      8
T09059
notch4 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jan-2000
C:Accession: T09059
R:Rowen, L.; Mahairas, G.; Qin, S.; Ahern, M.E.; Dankers, C.; Lasky, S.; Loretz, C., et al.
submitted to the EMBL Data Library, October 1997
A:Description: Sequence of the mouse major histocompatibility locus class III region.
A:Reference number: Z16543
A:Accession: T09059
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1964 <ROW>
A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564947
C:Genetics:
A:Gene: notch4
A:Map position: 17
A:Introns: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1;
1679/3; 1729/1; 1761/3
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: receptor; signal transduction
F:514-545/Domain: EGF homology <EGF>

Query Match          20.2%; Score 62.5; DB 2; Length 1964;
Best Local Similarity 30.9%; Pred. No. 79;
Matches 17; Conservative 7; Mismatches 18; Indels 13; Gaps 3;

QY      5  CPGGINAMNTISY-----IDNOICOGKRLCNTGDPKCPENGSC--VPDS 50
        ||| :||| | | | | | | | | | | | | | | | |
Db       202 CPQGTSCHNTLASYGCLCPVGGEPPQKLRKGAC---PPSSCLNGGTQLVPEG 252

RESULT      9
S27985
cellulase (EC 3.2.1.4) precursor - Streptomyces reticuli
N:Alternate names: avicelase; endo-1,4-beta-glucanase
C:Species: Streptomyces reticuli
```

C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 22-Oct-1999  
 C:Accession: S27985; S21398  
 R:Schlichtemeier, A.; Walter, S.; Schroeder, J.; Moorman, M.; Schrempf, H.  
 Mol. Microbiol. 6, 3611-3621, 1992  
 A:Title: The gene encoding the cellulase (Avicelase) Cell from Streptomyces reticuli and  
 A:Reference number: S27985; MUID:93116600  
 A:Accession: S27985  
 A:Molecule type: DNA  
 A:Residues: 1-746 <SCCH>  
 A:Cross-references: EMBL:X65616; NID:9683711; PIDN:CAA46570.1; PID:9581728  
 C:Genetics:  
 A:Gene: cell  
 A:Start codon: GTG  
 C:Function:  
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
 A:Pathway: cellulose degradation  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F:1-29/Domain: signal sequence #status predicted <SIG>  
 F:30-746/Product: cellulase #status predicted <MAT>

Query Match 19.9%; Score 61.5; DB 2; Length 746;  
 Best Local Similarity 30.0%; Pred. No. 41;  
 Matches 15; Conservative 8; Mismatches 12; Indels 15; Gaps 2;

Oy 7 GGANMNTTSYDNOICQGGKNCNMNGPE-----MCPENGSCVPD 49  
 Db 350 GGATWELLSTY-----ERSLTARTGHPALDGTALPESGNKVPD 391

RESULT 10

sodium channel protein II - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 20-Aug-1999  
 C:Accession: B25019; S24804  
 R:Noda, M.; Ikeda, T.; Kayano, T.; Takeshima, H.; Kurasaki, M.; Takahashi, H.  
 Nature 320, 188-192, 1986  
 A:Title: Existence of distinct sodium channel messenger RNAs in rat brain.  
 A:Reference number: A93377; MUID:86146901  
 A:Accession: B25019  
 A:Molecule type: mRNA  
 A:Residues: 1-2005 <NOD>  
 A:Experimental source: Brain  
 R:Sarao, R.; Gupta, S.K.; Auld, V.J.; Dunn, R.J.  
 Submitted to the EMBL Data Library, August 1991  
 A:Description: Developmentally regulated RNA splicing of rat brain sodium channel mRNAs.  
 A:Reference number: S24803  
 A:Accession: S24804  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 183-188, 'D', 190-305 <SAR>  
 A:Cross-references: EMBL:X61149; NID:957074; PIDN:CAA3458.1; PID:957076  
 C:Superfamily: sodium channel protein  
 C:Keywords: duplication; ion transport; sodium channel; transmembrane protein; voltage-g

Query Match 19.9%; Score 61.5; DB 2; Length 2005;  
 Best Local Similarity 33.3%; Pred. No. 1e+02;  
 Matches 16; Conservative 8; Mismatches 17; Indels 7; Gaps 3;

Oy 9 INANMNTTSYDNO-----ICOGQKN--LCNNTGDEMPENGSCVPDG 50  
 Db 311 VNMFNW-DEYIEDSKSHFYLEGQNDALLCGNSDAGCPEYICVKAG 357

RESULT 11

A57278  
 fibrillin-2 precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 11-Jan-2000  
 C:Accession: A57278  
 R:Zhang, H.; Hu, W.; Ramirez, F.

J. Cell Biol. 129, 1165-1176, 1995  
 A:Title: Developmental expression of fibrillin genes suggests heterogeneity of extrac  
 A:Reference number: A57278; MUID:95263670  
 A:Accession: A57278  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2907 <ZHA>  
 A:Cross-references: GB:L39790; NID:9762830; PIDN:AAA74908.1; PID:9762831  
 C:Superfamily: unassigned EGF-related proteins; EGF homology  
 F:1239-1274/Domain: EGF homology <EGF1>  
 F:2488-2523/Domain: EGF homology <EGF>

Query Match 19.9%; Score 61.5; DB 2; Length 2907;  
 Best Local Similarity 40.4%; Pred. No. 1.5e+02;  
 Matches 19; Conservative 3; Mismatches 10; Indels 15; Gaps 4;

Oy 19 IDNQCQGO-----KNLCNNT-----GDP-EMCPENGSCVPDPG 52  
 Db 220 VNMQGQGLTGIVCTKTLCCATIGRAGHCPGCPAPQ--PCRPG 264

RESULT 12

sodium channel alpha subunit - human  
 C:Species: Homo sapiens (man)  
 C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Aug-1999  
 C:Accession: I54323  
 R:McClatchey, A.I.; Lin, C.S.; Wang, J.; Hoffman, E.P.; Rojars, C.; Gussella, J.F.  
 Hum. Mol. Genet. 1, 521-527, 1992  
 A:Title: The genomic structure of the human skeletal muscle sodium channel gene.  
 A:Reference number: I54323; MUID:93338444  
 A:Accession: I54323  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1835 <RSS>  
 A:Cross-references: GB:L01983; NID:9337992; PIDN:AAA75557.1; PID:9908809  
 C:Genetics:  
 A:Gene: GDB:SCN4A  
 A:Cross-references: GDB:125181; OMIM:170500  
 A:Map position: 17q23.1-17q25.3  
 A:Introns: 91/3; 131/2; 161/2; 204/2; 235/1; 346/1; 367/2; 414/3; 484/3; 536/1; 615/3  
 C:Superfamily: sodium channel protein  
 C:Keywords: duplication

Query Match 19.6%; Score 60.5; DB 2; Length 1835;  
 Best Local Similarity 24.1%; Pred. No. 1.2e+02;  
 Matches 14; Conservative 7; Mismatches 20; Indels 17; Gaps 2;

Oy 10 NANMNTTSYDNO-----ICOGQKN--LCNNTGDEMPENGSCVPDG 50  
 Db 322 DTWNASHASNATNDTFWDAYISDEGNFYFLEGSNDALLCGNSDAGCPEYIECTKTKG 379

RESULT 13

sodium channel alpha chain - human  
 C:Species: Homo sapiens (man)  
 C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 20-Aug-1999  
 C:Accession: J50648; A42099  
 R:Wang, J.; Rojars, C.V.; Zhou, J.; Schwartz, L.S.; Nicholas, H.; Hoffmann, E.P.  
 Biochem. Biophys. Res. Commun. 182, 794-801, 1992  
 A:Title: Sequence and genomic structure of the human adult skeletal muscle sodium cha  
 A:Reference number: J50648; MUID:92134303  
 A:Accession: J50648  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1836 <MAN>  
 A:Note: 861-asp was also found as the result of polymorphism  
 R:McClatchey, A.I.; Van den Bergh, P.; Pericak-Vance, M.A.; Raskind, W.; Verellen, C.  
 Cell 68, 769-774, 1992  
 A:Title: Temperature-sensitive mutations in the III-IV cytoplasmic loop region of the



Thu Mar 28 09:21:38 2002

us-09-726-348-2\_copy\_126\_177\_1.rpt



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:17:28 ; Search time 35.36 seconds

(Without alignments)  
112.021 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_126\_177

Perfect score: 52  
Sequence: 1 OHVNCPCGINAMNTITSYID.....NTGPEMCPENGSCVPDGP 52

Scoring table: OLIGO  
Gapop 60.0 , Gapept 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6208

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database : PIR\_68:\*\*

1: p1r1:\*\*

2: p1r2:\*\*

3: p1r3:\*\*

4: p1r4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5	9.6	25	2	S10850 alpha-amylase inh
2	5	9.6	28	2	A46690 stalloic acid-specif
3	4	7.7	10	2	A49581 stallokinin I yel
4	4	7.7	10	2	A48778 small nuclear ribo
5	4	7.7	12	2	C58502 58k bile and gallb
6	4	7.7	13	2	S33800 chaperone, TCPI-re
7	4	7.7	16	2	A60839 neurokinin A homol
8	4	7.7	18	2	B24867 scyllorhizin II -
9	4	7.7	19	2	S54848 succinyl-CoA synth
10	4	7.7	19	2	A34233 trehalase inhibito
11	4	7.7	20	2	PA0022 protein QAI0011 -
12	4	7.7	21	2	S23361 protein-tyrosine k
13	4	7.7	21	2	B59325 probable bacteriop
14	4	7.7	23	2	B4304 ribosomal protein
15	4	7.7	24	2	S21511 incFVI protein - E
16	4	7.7	25	2	S17561 ubiquitin thiolest
17	4	7.7	27	2	S15989 streptomycin 3'-a
18	4	7.7	27	2	A44636 homeotic protein H
19	4	7.7	27	2	A84477 5S ribosomal RNA f
20	4	7.7	29	2	B43937 endo-1,4-beta-xyla
21	4	7.7	30	2	B54037 splicing regulator
22	3	5.8	4	2	I54357 schwannomin - mus
23	3	5.8	5	2	B23565 R-phycocerythrin al
24	3	5.8	6	2	A19780 transferrin - bovi
25	3	5.8	6	2	PT0709 T-cell receptor be
26	3	5.8	7	2	B61491 seed protein ws-5
27	3	5.8	7	2	H33098 180K exantigen -
28	3	5.8	7	2	A33098 244k exantigen -
29	3	5.8	7	2	PT0515 T-cell receptor be

30	3	5.8	8	2	PT0184 capsid protein VP-
31	3	5.8	8	2	A28719 thymic humoral fac
32	3	5.8	8	2	A28719 Ca2+-transporting
33	3	5.8	9	2	A14632 oxytocin - hippo
34	3	5.8	9	2	A91466 oxytocin - spotted
35	3	5.8	9	2	A92774 oxytocin - finback
36	3	5.8	9	2	A93147 oxytocin - Austral
37	3	5.8	9	2	A93408 oxytocin - rabbit
38	3	5.8	9	2	B90667 isotocin - common
39	3	5.8	9	2	B28495 conopressin S - co
40	3	5.8	9	2	A28495 conopressin G - co
41	3	5.8	9	2	S19329 sperm-activating p
42	3	5.8	9	2	B45020 probable minipolyp
43	3	5.8	9	2	S55696 phosphoenolpyruvat
44	3	5.8	9	2	PC2021 oxytocin-related p
45	3	5.8	9	2	A29477 diuretic neuropept

## ALIGNMENTS

RESULT 1  
S10850  
alpha-amylase inhibitor - durum wheat (fragment)  
N:Alternate names: glutenin low molecular weight chain  
C:Species: Triticum durum (durum wheat)  
C>Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
C:Accession: S10850  
R:Kobrehel, K.; Alary, R.  
J. Sci. Food Agric. 48, 441-452, 1989  
A:Title: Isolation and partial characterisation of two low molecular weight whe  
A:Reference number: S10849  
A:Accession: S10850  
A:Molecule type: protein  
A:Residues: 1-25 <KOB>  
C:Superfamily: wheat alpha-amylase inhibitor  
C:Keywords: alpha-amylase inhibitor

Query Match 9.6%; Score 5; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 GSCVP 48  
Db 2 GSCVP 6

RESULT 2  
A46690  
stalloic acid-specific O-acetyltransferase small subunit - rat (fragment)  
N:Alternate names: LSE small subunit  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 03-Mar-1995  
C:Accession: A46690  
R:Butor, C.; Higa, H.H.; Yarkl, A.  
J. Biol. Chem. 268, 10207-10213, 1993  
A:Title: Structural, immunological, and biosynthetic studies of a stalloic acid-specifi  
A:Reference number: A46690; MUID:93252902  
A:Accession: A46690  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-28 <BUT>  
A:Experimental source: liver  
A>Note: sequence extracted from NCBI backbone (NCBIP:131239)

Query Match 9.6%; Score 5; DB 2; Length 28;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 17 SYIDN 21  
|||||

Db 7 SYDN 11

RESULT 3

A:Accession: A49581

C:Species: Aedes aegypti (yellow fever mosquito)

C:Date: 06-Oct-1994 #sequence\_revision 18-Nov-1994 #text\_change 18-Aug-2000

C:Accession: A49581

R:Chapman, D.E.; Ribeiro, J.M.

Proc. Natl. Acad. Sci. U.S.A. 91, 138-142, 1994

A:Title: Salivary gland I and II: vasodilatory tachykinins from the yellow fever mosquito A

A:Reference number: A49581; MUID:94105119

A:Accession: A49581

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-10 <CH>

A:Experimental source: Rockefeller, salivary gland

A:Note: sequence extracted from NCBI backbone (NCBIP:141841)

C:Superfamily: unassigned animal peptides

Query Match 7.7%; Score 4; DB 2; Length 10;

Best Local Similarity 100.0%; Pred. No. 3.5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 NTGD 36

Db 1 NTGD 4

RESULT 4

small nuclear ribonucleoprotein E - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999

C:Accession: I48778

R:Fautsch, M.P.; Thompson, M.A.; Holicky, E.L.; Schultz, P.J.; Halleck, J.B.; Wieben, E.

Genomics 14, 883-890, 1992

A:Title: Conservation of coding and transcriptional control sequences within the snRNP E

A:Reference number: A44368; MUID:93122798

A:Accession: I48778

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-10 <RES>

A:Cross-references: EMBL:X65703; NID:g312006; PIDN:CAA46625.1; PID:g312007

Query Match 7.7%; Score 4; DB 2; Length 10;

Best Local Similarity 100.0%; Pred. No. 3.5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 OGOK 28

Db 6 OGOK 9

RESULT 5

CS8502

58k bile and gallbladder stone protein - unidentified bacterium (fragment)

C:Species: unidentified bacterium

C:Date: 07-Feb-1997 #sequence\_revision 07-Feb-1997 #text\_change 10-Jul-1998

C:Accession: C58502

R:Binette, J.P.; Binette, M.B.

submitted to the Protein Sequence Database, October 1996

A:Description: The proteins of kidney and gallbladder stones.

A:Reference number: A58501

A:Accession: C58502

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-12 <BIN>

A:Experimental source: human bile and gallbladder stones

Query Match 7.7%; Score 4; DB 2; Length 12;

Best Local Similarity 100.0%; Pred. No. 4.1e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 ENGS 45

Db 6 ENGS 9

RESULT 6

S33800

chaperone, TCPI-related - oat

C:Species: Avena sativa (oat)

C:Date: 02-Dec-1993 #sequence\_revision 27-Feb-1997 #text\_change 17-Mar-1999

C:Accession: S33800

R:Mummert, E.; Grimm, R.; Speth, V.; Ekerskorn, C.; Schiltz, E.; Gatenby, A.A.; Schla

Nature 363, 644-648, 1993

A:Title: A TCPI-related molecular chaperone from plants refolds phytochrome to its ph

A:Reference number: S33800; MUID:93288140

A:Accession: S33800

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-13 <NUM>

Query Match 7.7%; Score 4; DB 2; Length 13;

Best Local Similarity 100.0%; Pred. No. 4.3e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 DGPG 52

Db 5 DGPG 8

RESULT 7

A60839

neurokinin A homolog - marbled electric ray

N:Alternate names: des-Ser(1), Pro(2) scyllorhinin II

C:Species: Torpedo marmorata (marbled electric ray)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C:Accession: A60839

R:Conlon, J.M.; Thim, L.

Gen. Comp. Endocrinol. 71, 383-388, 1988

A:Title: Isolation of the tachykinin, Des[Ser(1)Pro(2)] scyllorhinin II from the inte

A:Reference number: A60839; MUID:89053024

A:Accession: A60839

A:Molecule type: protein

A:Residues: 1-16 <CON>

C:Keywords: amidated carboxyl end; neuropeptide; tachykinin

F:16/Modified site: amidated carboxyl end (Met) #status experimental

Query Match 7.7%; Score 4; DB 2; Length 16;

Best Local Similarity 100.0%; Pred. No. 5.2e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PDGP 51

Db 6 PDGP 9

RESULT 8

B24867

scyllorhinin II - smaller spotted catshark

C:Species: Scyliorhinus canicula (smaller spotted catshark, smaller spotted dogfish)

C:Date: 02-Jun-1988 #sequence\_revision 02-Jun-1988 #text\_change 08-Dec-1995

C:Accession: B24867

R:Conlon, J.M.; Deacon, C.F.; O'Toole, L.; Thim, L.

FEBS Lett. 200, 111-116, 1986

A:Title: Scyllorhinin I and II: two novel tachykinins from dogfish gut.

A:Reference number: A91359; MUID:86192829

A:Accession: B24867

A:Molecule type: protein  
 A:Residues: 1-18 <CON>  
 C:Keywords: amidated carboxyl end; neuropeptide  
 F:18/Modified site: amidated carboxyl end (Met) #status experimental

Query Match 7.7%; Score 4; DB 2; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PDGP 51  
 ||||  
 DB 8 PDGP 11

RESULT 9  
 S54848  
 succinyl-CoA synthetase (EC 6.2.1.-) alpha subunit - Pseudomonas aeruginosa (fragment)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 17-Jul-1998  
 C:Accession: S54848  
 R:Liao, X.; Lightfoot, J.; Charlebois, I.; Ouellet, C.; Morency, M.J.; Dewar, K.; Stienne  
 submitted to the EMBL Data Library, January 1995  
 A:Description: Physical mapping of 44 loci including accA, alme, ampC, ampR, arcA, aroK,  
 prp, pbbB, pbpc, phes, phoA, phoB, phoS, ponA, proS, pyoJ, qin, rpoB, rpoH, rpsB, sodB,  
 A:Reference number: S54841  
 A:Accession: S54848  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-19 <LIA>  
 A:Cross-references: EMBL:X84052  
 C:Genetics:  
 A:Gene: sucD  
 C:Superfamily: succinate-CoA ligase (ADP-forming) alpha chain  
 C:Keywords: acid-cholel ligase; ATP; phosphonitidine; phosphoprotein; tricarboxylic acid

Query Match 7.7%; Score 4; DB 2; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 6.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 ICQG 26  
 ||||  
 DB 12 ICQG 15

RESULT 10  
 A34233  
 trehalase inhibitor - American cockroach (fragment)  
 C:Species: Periplaneta americana (American cockroach)  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 18-Jun-1993  
 C:Accession: A34233  
 R:Hayakawa, Y.; Jahagirdar, A.P.; Yaguchi, M.; Downer, R.G.H.  
 J. Biol. Chem. 264, 16165-16169, 1989  
 A:Title: Purification and characterization of trehalase inhibitor from hemolymph of the  
 A:Reference number: A34233; MUID:89380218  
 A:Accession: A34233  
 A:Molecule type: protein  
 A:Residues: 1-19 <HAY>

Query Match 7.7%; Score 4; DB 2; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 6.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VPDG 50  
 ||||  
 DB 12 VPDG 15

RESULT 11  
 PA0022  
 protein OA100011 - Arabidopsis thaliana (fragment)

C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 30-Jun-1992 #sequence\_revision 06-Jan-1995 #text\_change 06-Jun-1997  
 C:Accession: PA0022  
 R:Kamo, M.; Kawakami, T.; Miyatake, N.; Tsugita, A.  
 submitted to JIPID, July 1994  
 A:Description: Separation and characterization of Arabidopsis proteins by two-dimensi  
 A:Reference number: PA0001  
 A:Accession: PA0022  
 A:Molecule type: protein  
 A:Residues: 1-20 <KAM>  
 A:Experimental source: callus

Query Match 7.7%; Score 4; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 6.3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 47 VPDG 50  
 ||||  
 DB 9 VPDG 12

RESULT 12  
 S23361  
 protein-tyrosine kinase (EC 2.7.1.112) eek - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 04-Feb-2000  
 C:Accession: S23361  
 R:Chan, J.; Watt, V.M.  
 Oncogene 6, 1057-1061, 1991  
 A:Title: eek and erk, new members of the eph subclass of receptor protein-tyrosine ki  
 A:Reference number: S23361; MUID:91296384  
 A:Accession: S23361  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-21 <CHNA>  
 A:Cross-references: EMBL:X59291  
 C:Genetics:  
 A:Gene: GDB:EEK  
 A:Cross-references: GDB:125195; OMIM:176945  
 A:Map position: 1pter-1qter  
 C:Superfamily: protein-tyrosine kinase, receptor type eph; fibronectin type III repea  
 C:Keywords: ATP; phosphotransferase; transmembrane protein; tyrosine-specific protein

Query Match 7.7%; Score 4; DB 2; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 ENGS 45  
 ||||  
 DB 12 ENGS 15

RESULT 13  
 B59325  
 probable bacteriophage receptor BactB [imported] - Bacteroides fragilis (fragment)  
 C:Species: Bacteroides fragilis  
 C:Date: 28-Jul-2000 #sequence\_revision 28-Jul-2000 #text\_change 28-Jul-2000  
 C:Accession: B59325  
 R:Frias-Lopez, J.  
 submitted to the Protein Sequence Database, July 2000  
 A:Description: Identification of cell wall proteins of Bacteroides fragilis to which  
 A:Reference number: A59325  
 A:Accession: B59325  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-21 <FRL>  
 A:Experimental source: strain ATCC 51477  
 A:Note: putative receptor for bacteriophage B40-8; residue 6 may be Val, 7 may be Gly

Query Match 7.7%; Score 4; DB 2; Length 21;

Best Local Similarity 100.0%; Pred. No. 6.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 GINA 11  
||||  
Db 5 GINA 8

## RESULT 14

B24304  
ribosomal protein C [validated] - Haloarcula marismortui (fragment)  
C:Species: Haloarcula marismortui  
C:Date: 19-May-1989 #sequence\_revision 19-May-1989 #text\_change 21-Jul-2000  
C:Accession: B24304  
R:Shoham, M.; Dijk, J.; Reinhardt, R.; Wittmann-Liebold, B.  
FEBS Lett. 204, 323-330, 1986  
A:Title: Purification and characterization of ribosomal proteins from the 30 S subunit  
A:Reference number: A24304  
A:Accession: B24304  
A:Molecule type: protein  
A:Residues: 1-23 <SHO>  
A:Note: the source is designated as Haloaracterium marismortui  
C:Keywords: protein biosynthesis; ribosome

Query Match 7.7%; Score 4; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 7.2e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 DPEM 39  
||||  
Db 17 DPEM 20

## RESULT 15

S21511  
incFVI protein - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999  
C:Accession: S21511  
R:Andres, L.; Rodriguez, O.  
submitted to the EMBL Data Library, September 1990  
A:Reference number: S21510  
A:Accession: S21511  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-24 <AND>  
A:Cross-references: EMBL:X55895; NID:g42544; PIDN:CAA39381.1; PID:e33255; PID:g1333753

Query Match 7.7%; Score 4; DB 2; Length 24;  
Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 DPG 52  
||||  
Db 5 DPG 8

Search completed: March 28, 2002, 09:17:28  
Job time: 382 sec





```

RESULT      4
A24420
notch protein - fruit fly (Drosophila melanogaster)
N:Alternate names: neurogenic repetitive locus protein
C:Species: Drosophila melanogaster
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A24420; A24768; S09358; A05267
R:Kidd, S.; Kelley, M.R.; Young, M.W.
Mol. Cell. Biol. 6, 3094-3108, 1986
A:Reference number: A24420; MUID:87064624
A:Accession: A24420
A:Molecule type: DNA
A:Residues: 1-2703 <RID>
A:CROSS-references: GB:K03508; MID:g157991; PID:AAA28725.1; PID:g157993
R:Wharton, K.A.; Johanson, K.M.; Xu, T.; Artavanis-Tsakonas, S.
Cell 43, 567-581, 1985
A:Reference number: A24768; MUID:86079539
A:Accession: A24768
A:Molecule type: mRNA
A:Residues: 1-48, 'I', '50-118, 'R', '120-230, 'I', '232-256, 'N', '258-266, 'A', '268-872, 'R', '874-958
A>Note: The authors translated the codon ATC for residue 49 as Thr, ATT for residue 2044
R:Tautz, D.
Nucleic Acids Res. 17, 6463-6471, 1989
A:File: Hypervariability of simple sequences as a general source for polymorphic DNA m...
A:Reference number: S09358; MUID:89385974
A:Accession: S09358
A:Molecule type: DNA
A:Residues: 2505-2551, 'QOQQ', '2552-2576, 'E', '2578-2604 <TAU>
R:Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.
Cell 40, 55-62, 1985
A:File: opa: a novel family of transcribed repeats shared by the Notch locus and other
A:Reference number: A05267; MUID:85099329
A:Accession: A05267
A:Molecule type: DNA
A:Residues: 2504-2576, 'E', '2578-2611 <WHA2>
C:Genetics:
A:Gene: notch; opa
A:CROSS-references: FlyBase:FBN0004647
A:Map position: 8, 96-9.36
A:Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
C:Keywords: differential; tandem repeat; transmembrane protein
F:27-43/Domain: transmembrane #status predicted <TMH1>
F:297-328/Domain: EGF homology <EGX1>
F:530-561/Domain: EGF homology <EGF1>
F:568-599/Domain: EGF homology <EGF>
F:988-1019/Domain: EGF homology <EGX2>
F:1064-1095/Domain: EGF homology <EGF3>
F:1187-1218/Domain: EGF homology <EGX3>
F:1746-1762/Domain: transmembrane #status predicted <TMW2>
F:1950-1982/Domain: ankyrin repeat homology <AN1>
F:1983-2015/Domain: ankyrin repeat homology <AN2>
F:1988-2004/Domain: transmembrane #status predicted <TMW3>
F:2017-2049/Domain: ankyrin repeat homology <AN3>
F:2050-2082/Domain: ankyrin repeat homology <AN4>
F:2083-2115/Domain: ankyrin repeat homology <AN5>
F:2538-2568/Region: glutamine-rich
F:2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match          8.4%; Score 105; DB 1; Length 2703;
Best Local Similarity 24.6%; Pred. No. 1.1;
Matches 47; Conservative 17; Mismatches 59; Indels 68; Gaps 10;

QY      27 RALALPEICITCGPGSVONLSKYAF-----YCK-----TTREIMLNAR 63
       :|::|:| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
DB      1147 KGLSRLDLNNCTCKDYGNSHVCYSQGAGSYCQKEIDECSOPCPNGGTRDDLIGAVE 1206
       ::||::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |
QY      64 CCLINKGITLIGLDNLNCSLEDDPGPNFHQAHTVVIIDLIQANPLK--GDLANFRGEFTLOLT 121
       || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1207 GCQRQ-----GRGGNCCELN-----IDDCANPCQNGSGTCHDRVMNTS---- 1244
       || | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      122 LILPQHVNCPGGINAMNWTITSYIIDNOICOGKNLCNNGTEDPMECBPNSSCVDPGPELLQC 181

```

```

Db      1245  -----CCCPG-----TMT-----ICELNKDKCK-----PGACHNNNSCI-DRYVGGEFC 1283
QY      182  VCADEGFHGKYC 192
           || || |
Db      1284  VCGPGFVGARC 1294

RESULT   5
SI6148
gene serrate protein precursor - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 31-Dec-1991 #sequence revision 02-Aug-1994 #text_change 17-Nov-2000
C:Accession: SI6148; SI6878; A3666
R:Thomas, U.; Speicher, S.A.; Knust, E.
A:Title: The Drosophila gene Serrate encodes an EGF-like transmembrane protein with a
A:Accession: SI6148
A:Molecule type: mRNA
A:Residues: 1-1408 <TH01>
A:Cross-references: EMBL:X56811
R:Thomas, U.
submitted to the EMBL Data Library, November 1990
A:Reference number: SI6878
A:Accession: SI6878
A:Molecule type: mRNA
A:Residues: 1-1351, 'T', 1353-1408 <TH02>
A:Cross-references: EMBL:X56811; NID:98564
R:Fleming, R.J.; Scottgale, T.N.; Diederich, R.J.; Artavanis-Tsakonas, S.
Gene Dev. 4, 2188-2201, 1990
A:Title: The gene Serrate encodes a putative EGF-like transmembrane protein essential
A:Reference number: A3666; MUID:91099666
A:Accession: A3666
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-15, 20-26, 'A', 28-1408 <FE>
A:Cross-references: GB:M35759; NID:9158605; PID:9158606
C:Genetics:
A:Gene: FlyBase:Ser
A:Cross-references: FlyBase:FBgn0004197
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: glycoprotein; transmembrane protein
F:1-84/Domain: signal sequence #status predicted <SIG>
F:85-1408/Product: gene serrate protein #status predicted <MAT>
F:85-1221/Domain: extracellular #status predicted <EXT>
F:283-316/Domain: EGF homology <EG01>
F:319-348/Domain: EGF homology <EG02>
F:355-388/Domain: EGF homology <EG03>
F:395-488/Domain: EGF homology #status atypical <EG04>
F:495-526/Domain: EGF homology <EG05>
F:533-608/Domain: EGF homology #status atypical <EG06>
F:615-645/Domain: EGF homology <EG07>
F:652-683/Domain: EGF homology <EG08>
F:690-720/Domain: EGF homology <EG09>
F:727-796/Domain: EGF homology #status atypical <EG10>
F:803-834/Domain: EGF homology <EG11>
F:841-876/Domain: EGF homology <EG12>
F:883-914/Domain: EGF homology <EG13>
F:921-952/Domain: EGF homology <EG14>
F:997-1000/Region: cysteine-rich
F:1222-1246/Domain: transmembrane #status predicted <TM1>
F:1247-1408/Domain: intracellular #status predicted <INT>
F:152,196,247,331,412,452,558,739,965,977,1004,1030,1150/binding site: carbohydrate (

```

```

C:Keywords: acceptor; signal transduction
F:514-545/Domain: EGF homology <EGF>

Query Match      8.1%; Score 101; DB 2; Length 1964;
Best Local Similarity 21.1%; Pred. No. 1.8;
Matches 48; Conservative 24; Mismatches 75; Indels 80; Gaps 11.

QY      35 CTCCPGSVQMLSKVARYCTKTRFLMLHANC-----CLNQKGTILGL-----DLQ 78
          ||      ||      ||      ||      ||      ||      ||      ||
Db       750 CTCLP-----SHTGRHCQTAVDHCVSASCLNGTGVKPGTFFCLCATGFGHCEKT 803
          ||      ||      ||      ||      ||      ||      ||      ||
QY      79 NCSLEDP-----GPNFHOAHTTVIID-----LQANPLK 106
          ||      ||      ||      ||      ||      ||      ||      ||
Db       804 NPSCADSPCKNKATQODTRGARCICLSPGYTSSCOTLIDLCAKPCPHARCICQSPSF 863
          ||      ||      ||      ||      ||      ||      ||      ||
QY      107 GDLANTRGFT-----QLQTLILPQAHV---CPGGINAMWTTTSY-----IDNQ 147
          ||      ||      ||      ||      ||      ||      ||      ||
Db       864 QCLC--LQGTWGLADCPFLSCQKAAMSGCIEISGLQNGSLCIDPTGSSYFCRCPFGQK 921
          ||      ||      ||      ||      ||      ||      ||      ||
QY      148 ICQGGKMLCNNTGDPKEMCPENGSCVPADGGGLLQCYCAGGFFGYGCMR 194
          ||      ||      ||      ||      ||      ||      ||      ||
Db       922 LQDQNVNPC---EPNPSCHNGSTCVAPQPSGYV-CQCAAGYEGYGQMSK 963
          ||      ||      ||      ||      ||      ||      ||      ||

RESULT      8
A55567
Fibrillin I - bovine
Species: Bos primigenius taurus (cattle)

```

```

A>Title: Sequence of the coding region of the bovine fibrillin cdna and local
A:Reference number: A55567; MUID:95137597
A:Accession: A55567
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2871 <Tili>
A:Cross-references: GB:I28748; NID:g508427; PIDN:AAA74122.1; PID:g508428
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:I201-1236/Domain: EGF homology <Eef>

Query Match      8 0%: Score 99.5; DB 2; Length 2871;
Best Local Similarity 21.2%; Pred. No.3.5; Indels 79; Gaps 11.
Matches 46; Conservative 19; Mismatches 73;

QY      1  MAPHPSGLTITVPMAALLALGVERALALPEICT--QC---PGSVONLSKYA----- 49
      Db 2389  LCPHGRGFMN-----GADIDCEVYIHIDVARNECYNDGSGYHCICKTKGYTPDIT 2438
      QY      50  -----FYCKTTEELMLHARCCLNOKGTLLGIIDLQNC-SLEDPGEFHF 90
      Db 2439  GTACVLDNECNAQPKPCNTICKNTE----GSYQSCSPRGYIILQDGRSKCDLDECATKQH 2494
      QY      91  QAHRTVITLDQANPLKGLDANTFRGFTQLTQTLILPQHYNCGCGGINMANNTTSTYDNOIC 150
      Db 2495  NCOFLCY-----NTIGSFT-----CKCPFTQHH--TACIDNNECH 2529
      QY      151  GQKNLCNNITGDPMPCEPENGSCVPDDPGILGQCYCADGF 187
      Db 2530  SDINLCGSKG--ICQNT-----PGSFCECGQNGF 2556

RESULT      9
165967
disintegrin-like metalloproteinase (EC 3.4.24.-), splice form 2 - human
N:Alternate names: ADAM11; KDC
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 17-Mar-2000 #text_change 26-May-2000
C:Accession: I65967; S38539
C:Kataagiri, T.; Harada, Y.; Eml, M.; Nakamura, Y.

```



Cytogenet. Cell Genet. 68, 39-44, 1995  
 A:Title: Human metalloproteinase/disintegrin-like (MDC) gene: exon-intron organization and  
 A:Reference number: 152965; MUID:95044425  
 A:Accession: I65967  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 7-670 <KAT>  
 A:Cross-references: GB:D31872; NID:9505090; PIDN:BA06670.1; PID:d1007242; PID:983683  
 R:Enl, M.; Katagiri, T.; Harada, Y.; Saito, H.; Inazawa, J.; Ito, I.; Kasumi, F.; Nakamu  
 Nature Genet. 5, 151-157, 1993  
 A:Title: A novel metalloproteinase/disintegrin-like gene at 17q21.3 is somatically rearran  
 A:Reference number: S38539; MUID:94073190  
 A:Accession: S38539  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-225, 'N', 227-495, 'POGGRVNLPRLCQHMSSARPGGRQ' <EMT>  
 A:Cross-references: GB:D17390; NID:9452188; PIDN:BA04213.1; PID:d1004732; PID:9484255  
 C:Comment: For an alternative splice form, see PIR:S38539.  
 C:Genetics:  
 A:Gene: GDB:MDC; ADAM11  
 A:Cross-references: GDB:230267; OMIM:155120  
 A:Map position: 17q21.3-17q21.3  
 A:Introns: 28/3; 57/2; 82/3; 105/1; 127/3; 152/3; 176/3; 232/2; 260/3; 291/1; 308/2; 341  
 C:Superfamily: disintegrin homology  
 C:Keywords: alternative splicing; hydrolase; metalloproteinase  
 F:344-427/Domain: disintegrin homology <DIS>

Query Match 8.0%; Score 99; DB 2; Length 670;  
 Best Local Similarity 21.4%; Pred. No. 0.87;  
 Matches 55; Conservative 32; Mismatches 90; Indels 80; Gaps 14;

QY 31 LPEICT-----OCPSYONLSKVAFYCK-----TTRE-----IMLHA---RCC-- 65  
 Db 413 IAEICTGDSOCPRNIAKLD--GYCHBQRCYCGCKTRDRQCYLWKAHAADRCYCE 470  
 QY 66 -LNQGT-----ILGLDNLCSLEDPGFNFQAHNTVILDLQANPLKGLDANTFRGFTQ 118  
 Db 471 KLVNVEGRGSCGRKSGWVCSKQDYLGP-----LLCVNISCAPRLGLDVGDISSVTF 525  
 QY 119 LQTLILPQHVNCPG--INMANTTSTYID-----NOICOGKNL-----CNNTG 160  
 Db 526 YHQ--GNELDRCRGHVLQDAGSDLSYEDGTACGPNMLCLDHRCPLPASAFNESTCPGSG 582  
 QY 161 DPEKPCNGSCVPDPGLDCCVACDGFHYKC-----MNGSGSLMNF 204  
 Db 583 ERRISSHGVCNNG--KTCQPDWTGKDCSINHPLPTSPPTGETERYKGPSTNITII 638  
 QY 205 GILGATLVSILLMAT 221  
 Db 639 GSINGAVLVAIVLGGT 655

RESULT 10  
 B66465  
 Probable Protein kinase [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: B66465  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultzar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzalli,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: B66465  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-1064 <STO>  
 A:Cross-references: GB:AE005172; NID:g10086466; PIDN:AAG12526.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1

Query Match 8.0%; Score 99; DB 2; Length 1064;  
 Best Local Similarity 26.4%; Pred. No. 1.4;  
 Matches 37; Conservative 14; Mismatches 51; Indels 38; Gaps 4;

QY 37 QCPGVSQNLKVAFYCKTTRMLNARCCNLQNGKGLGLDNLCSLEDPGFNFQAHNTV 96  
 Db 563 QIPKSIKLN-----OKTLILDLSTNSLSGEIPELQVTSILFI 600  
 QY 97 IIDLANPLKGLDANTFRGFTQLOTLILPQHVNCPSGINAMTITSYIDNOICOGKMLC 156  
 Db 601 NIDLSTNTTGTGNLPFRFSULTQSLDLSN-SLHGDIKVLGLSLASLANS-----C 653  
 QY 157 NNTGPEMCPENGSCVPDGP 176  
 Db 654 NNFGSP-----IPSTP 664

RESULT 11  
 T22674  
 hypothetical protein F54F3.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T22674  
 R:Pericy, C.; Lloyd, C.  
 submitted to the EMBL Data Library, September 1996  
 A:Reference number: Z19598  
 A:Accession: T22674  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1584 <WIL>  
 A:Cross-references: EMBL:Z79696; PIDN:CA01972.1; GSPDB:GN00023; CESP:F54F3.1  
 A:Experimental source: clone F54F3  
 C:Genetics:  
 A:Gene: CESP:F54F3.1  
 A:Map position: 5  
 A:Introns: 35/3; 157/3; 213/2; 257/1; 357/1; 440/3; 545/1; 854/3; 896/1; 944/1; 1001/

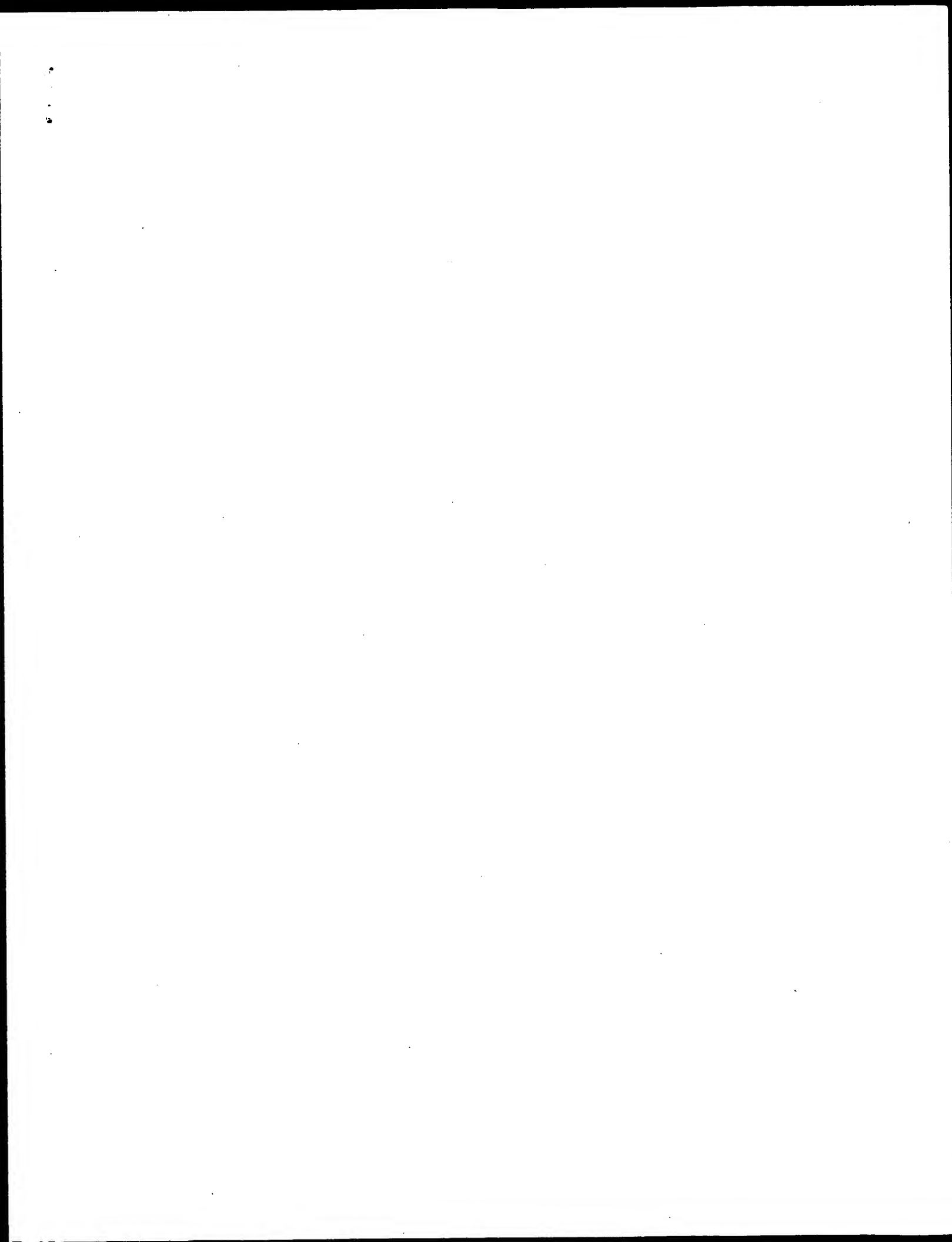
Query Match 7.9%; Score 98.5; DB 2; Length 1584;  
 Best Local Similarity 27.1%; Pred. No. 2.3;  
 Matches 35; Conservative 15; Mismatches 50; Indels 29; Gaps 7;

QY 79 NCSLEDPGFNFQAHNTVILDLQANPLK-----GDLANTFRGFTQLOTLILPQHVNC 130  
 Db 1156 NCNLE--PRICHANACVHHHTNAVEICIKPGSSGD-----GYTKCDVLETFRCTNC 1206  
 QY 131 PGGINAM--NTITSYIDNOIC-----OGKRLCNNTG---DPEMCPENGSCVPDGPGLQ 180  
 Db 1207 --SIHAYCAQNPFGAYQCCKNAGYNGHGLCVSMSSCLDRSLDENADVCPGEGAHYV 1264  
 QY 181 CVCADGPHG 189  
 Db 1265 CNCHYGHG 1273

RESULT 12  
 S42612  
 transmembrane protein precursor - zebra fish  
 C:Species: Brachydanio rerio (zebra fish)  
 C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Sep-1999  
 C:Accession: S42612  
 R:Bierkamp, C.; Campos-Ortega, J.A.  
 Mech. Dev. 43, 87-100, 1993  
 A:Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its patter  
 A:Reference number: S42612; MUID:94128602  
 A:Accession: S42612  
 A:Status: preliminary







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:11:06 ; Search time 35.36 Seconds

(without alignments)  
493.325 Million cell updates/sec

Title: US-09-726-348-2

Perfect score: 229

Sequence: 1 MAPHGPGSLTTLVPMMAALL.....TTLVSILMATORAKAKTS 229

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size: 0

Total number of hits satisfying chosen parameters: 6208

Minimum DB seq length: 0

Maximum DB seq length: 30

Post-processing: Listing first 45 summaries

Database:

1: PIR\_68:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6	2.6	30	2 I68109	interferon alpha-W
2	5	2.2	15	2 S36891	ribosomal protein
3	5	2.2	20	2 A60822	cytochrome P450 PB
4	5	2.2	20	2 A36016	granulocyte inhibi
5	5	2.2	22	2 D30609	Ig kappa chain V-I
6	5	2.2	23	2 I55406	nicotinic acetylch
7	5	2.2	23	2 B37843	vrg 18 protein - B
8	5	2.2	24	2 B30609	Ig kappa chain V-I
9	5	2.2	24	2 P00677	photosystem I 9.0K
10	5	2.2	24	2 S10681	probable 7-ethoxyc
11	5	2.2	24	2 S10682	cytochrome P450 is
12	5	2.2	25	2 PC1221	alanine racemase (
13	5	2.2	25	2 S10850	alpha-amylase inh
14	5	2.2	26	2 A32806	transposase - Esch
15	5	2.2	26	2 B44107	cytochrome P450 NR
16	5	2.2	26	2 A61056	aminopyrine N-deme
17	5	2.2	26	2 S05414	steryl-sulfatase (
18	5	2.2	27	2 I46566	antileukoproteinas
19	5	2.2	27	2 S55802	stp protein (Baker
20	5	2.2	27	2 S55801	stp protein (Baker
21	5	2.2	28	2 S49924	stp protein (Baker
22	5	2.2	28	2 S29135	aminopyrine N-deme
23	5	2.2	28	2 S29136	aminopyrine N-deme
24	5	2.2	28	2 A46690	sialic acid-specif
25	5	2.2	29	2 S28174	heat-shock protein
26	5	2.2	30	2 S21814	H+-transporting AT
27	5	2.2	30	2 E45095	phosphorylase I
28	5	2.2	30	2 S12902	diphtheria toxin f
29	4	1.7	9	2 S07241	litorin - Rohde's

30	4	1.7	10	1	ECT10AM
31	4	1.7	10	2	A49581
32	4	1.7	10	2	I48778
33	4	1.7	11	2	A33917
34	4	1.7	11	2	F33098
35	4	1.7	12	2	S25485
36	4	1.7	12	2	C58502
37	4	1.7	12	2	A26093
38	4	1.7	12	2	A56878
39	4	1.7	12	2	PH1635
40	4	1.7	12	2	PH1611
41	4	1.7	13	2	A38929
42	4	1.7	13	2	A60336
43	4	1.7	13	2	S33800
44	4	1.7	13	2	B61458
45	4	1.7	14	2	PA0109

#### ALIGNMENTS

RESULT 1  
I68109  
interferon alpha-WA precursor - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 04-Oct-1996 #sequence\_revision 04-Oct-1996 #text\_change 16-Jul-1999  
C:Accession: I68109  
R:Fuks, M.; Hendrix, L.C.; Bollon, A.P.  
Gene 32, 135-140, 1984  
A>Title: Pseudogene IFN-alpha L: removal of the stop codon in the signal sequence per  
A:Reference number: I53988; MUID:85155470  
A:Accession: I68109  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-30 <RES>  
A:Cross-references: GB:M10201; MUID:9184654; PIDN:AAA52732.1; PID:9184655  
C:Genetics:  
A:Gene: IFNA  
C:Superfamily: Interferon alpha

Query Match 2.6%; Score 6; DB 2; Length 30;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 197 SEFLM 202  
DB 4 SFSILM 9  
RESULT 2  
S36891  
ribosomal protein - Mycobacterium bovis (fragment)  
C:Species: Mycobacterium bovis  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Jan-1995  
C:Accession: S36891  
R:ohara, N.; Kimura, M.; Higashi, Y.; Yamada, T.  
FEBS Lett. 331, 9-14, 1993  
A>Title: Isolation and amino acid sequence of the 30S ribosomal protein S19 from Myco  
A:Reference number: S36887; MUID:94009653  
A:Accession: S36891  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-15 <OHAS>

Query Match 2.2%; Score 5; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 3; 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 223 RRRKX 227  
DB 10 RRRKX 14

RESULT 3  
cytochrome P450 PB-3a - rat (fragment)  
A:Reference number: A30601; MUID:89215279  
A:Accession: D30609  
A:Status: preliminary  
A:Molecule type: protein  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 2.2%; Score 5; DB 2; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 TLSVS 215  
DB 10 TLSVS 14

RESULT 4  
A36016  
granulocyte inhibitory protein - human  
C:Species: Homo sapiens (man)  
C:Date: 11-Jan-1991 #sequence\_revision 11-Jan-1991 #text\_change 30-May-1997  
C:Accession: A36016  
R:Hoerl, W.H.; Haeg-Weber, M.; Georgopoulos, A.; Block, L.H.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6353-6357, 1990  
A:Title: Physicochemical characterization of a polypeptide present in uremic serum that  
A:Reference number: A36016; MUID:90349614  
A:Accession: A36016  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-20 <H0E>  
C:Superfamily: immunoglobulin V region; immunoglobulin homology

Query Match 2.2%; Score 5; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LLLAL 23  
DB 7 LLLAL 11

RESULT 5  
D30609  
Ig kappa chain V-III regions (Jon and Mit) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 29-Jun-1989 #sequence\_revision 29-Jun-1989 #text\_change 30-May-1997  
C:Accession: D30609  
R:Goni, F.R.; Chen, P.P.; McGinnis, D.; Arjonilla, M.L.; Fernandez, J.; Carson, D.; Solc  
J. Immunol. 142, 3158-3163, 1989

Query Match 2.2%; Score 5; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 TLSVS 215  
DB 10 TLSVS 14

A:Title: Structural and idiotype characterization of the L chains of human IgM autoa  
A:Reference number: A30601; MUID:89215279  
A:Accession: D30609  
A:Status: preliminary  
A:Molecule type: protein  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotetramer; immunoglobulin

Query Match 2.2%; Score 5; DB 2; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 TLSVS 215  
DB 10 TLSVS 14

RESULT 6  
155406  
nicotinic acetylcholine receptor - Rattus leucopus (fragment)  
C:Species: Rattus leucopus  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Aug-1999  
C:Accession: 155406  
R:Walke, W.; Staple, J.; Adams, L.; Gneagy, M.; Chahine, K.; Goldman, D.  
J. Biol. Chem. 269, 19447-19456, 1994  
A:Title: Calcium-dependent regulation of rat and chick muscle nicotinic acetylcholine  
A:Reference number: 155406; MUID:94308229  
A:Accession: 155406  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-23 <RES>  
A:Cross-references: GB:LL9594; NID:g310180; PIDN:AAA21730.1; PID:g310181  
C:Genetics:  
A:Gene: nAChR  
C:Superfamily: acetylcholine receptor  
C:Keywords: neurotransmitter receptor

Query Match 2.2%; Score 5; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LLLAL 23  
DB 10 LLLAL 14

RESULT 7  
B37843  
vrg 18 protein - Bordetella pertussis (fragment)  
C:Species: Bordetella pertussis  
C:Date: 28-Jun-1991 #sequence\_revision 28-Jun-1991 #text\_change 17-Feb-1994  
C:Accession: B37843  
R:Beattie, D.T.; Knapp, S.; Mekalanos, J.J.  
J. Bacteriol. 172, 6997-7004, 1990  
A:Title: Evidence that modulation requires sequences downstream of the promoters of t  
A:Reference number: A37843; MUID:91072251  
A:Accession: B37843  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-23 <BEA>  
A:Cross-references: GB:M37228

Query Match 2.2%; Score 5; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AAALL 20  
DB 10 AAALL 14



Db	549	TGTG-TTGGAGCGATGGTTTCGCATGATNACAAGTAGTGCCACAGGGCTCCTTCACAATG	607
Qy	601	cttaagttccttcgggaattcttgaggcaaccactatccgtctccaatctgactttagcg	660
Db	608	CTTAAGTCTTCGCCGATTCCTGGAG-CACCAATCATTCGCTCCATTCCTCTGGGCG	666
Qy	661	acc 663 	
Db	667	ACC 669	
RESULT	15		
AL527415			
LOCUS	AL527415	980 bp	mRNA
DEFINITION	AL527415 LTI_NFL003_NBC3 Homo sapiens cDNA clone CSDDC021Y02 5 prime, mRNA sequence.	EST	13-FEB-2001
ACCESSION	AL527415		
VERSION	AL527415.1	GI:12790908	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Eumetaria; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Li W.B., Gruber C., Jesses J. and Polayes D.		
JOURNAL	Full-length cDNA libraries and normalization		
COMMENT	Unpublished (2001)		
	Contact: Genoscope		

**FEATURES**  
 BP 191 91006 EVRY cedex - France  
 Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Location/Qualifiers

```

source
1. 980
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS00C021YJ02"
/clone_lib="TRI-NFL003_NBC3"
/sex="male"
/issue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: PCMVSPORT 6; 1st strand cDNA was primed with a NotI-Oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Peng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : liang@lifestechn.com URL : http://fulllength.invitrogen.com"
218 a 269 c 294 g 195 t 4 others
BASE COUNT
ORIGIN

```

Query Match	91.08;	Score 625.2;	DB 10;	Length 980;
Best Local Similarity	98.88;	Pred. No. 1.8e-174;		
Matches 648; Conservative	3;	Mismatches 3;	Indels 2;	Gaps 2;

QY	1	atggcgcctcaacgagcccgagggaagctcttaagacccttggtgcgccttgagctc	60
Db	326	ATGGCGCCTCACGGCCCGGGGAGTCTTACGACCTGTGGCCCTGGCCCTGCTC	385
QY	61	ctgcctctggagcgttggaaagagcctctgagctaccgcgaagatgaccacaatgtccagg	120
Db	386	CTGCGCTTGGGCGTGGAAAGGGCTCTTGCGCTACCCAGATATGACACCCATCTCCAGGG	445
QY	121	agcgtgcacaatttgtcaaaagtgccctttatgttaaaagcagcagagcctaagtctg	180
Db	446	AGCGTGTCAAAATTTGTCAAAAGTGGCTTTATTATTGTAAACGACACGAGGCTAAATGCTG	505
QY	181	catgcccgcttgctgcctgaatcgaagagcgacacatcttggagcttgatctccgaactgt	240

Db	506	CAAGCCCTTCTGGCTGATACAGAGGGACCACTTTGGGGCTGGATCTCCAGAACTGT	565
OY	241	tctctggaagaacctgyltccaaccttcatcaaggacataaccactgylcatcatagacctg	300
Db	566	TCTCTGAGAGAACCTGGTCCAAACTTTCATCAGGCACTACCACTGTBATTATGACTCG	625b
OY	301	caagcaaacacctcnaaagltgacttgygcaaacacttcagtgtccttactcaagctcag	360
Db	626	CAAGCAAAACCCCTCAAAGGTGACTTGCCCAACCTTCGGTGGCTTTACTCAGCTCAG	685
OY	361	actctgatactgcgccaacaacalgtcaactgtctcgygaggaatlaatgcctgyaactatc	420
Db	686	ACTCTGATACTGCCACAACATGTCACTGTCTCTGGAGGAATTAATGCTCGAATACTATC	745
OY	421	acctctatataagacaaccaaatctgttccaagggcnaaaagaaccttgcataaactgyg	480
Db	746	ACCTCTTATATAGCAACCAACCAATCTGTSAGGGCCAAAAGAACCTTTGCATTAACACTGG	805
OY	481	gaccccaagaatgltcctctggaagaatcgttctgttacctg- atggtccaggtccttttga	539
Db	806	GACCCAGAAAGTGTCTCGAAGAAAGGATCTTGTTACTCAATAGTGTCACAGTCTTTTGCA	865
OY	540	gtgtgtttgttcctgaatggtttccatlgatbaacaagtatgacgccaaggtcgtttcact	599
Db	866	GTTGTGTTTGCTATGATGTTTCCATGTGATCAAACTGTATGTCGCGCAGGGCTGTTCTACT	925
OY	600	gctatgttctctcggaattctcggaagccacaacactctacgttccatcttggctt	655
Db	926	GCTTATGTGCTTCGGGATTTCTGGAG- CAACACTCTTATGCGTCTCCATTTGCTTT	980

Search completed: March 28, 2002, 07:58:33  
Job time: 30365 sec



BASE COUNT 219 a 269 c 291 g 189 t 1 others  
 ORIGIN

Query Match 91.3%; Score 627; DB 10; Length 969;  
 Best Local Similarity 99.5%; Pred. No. 5.4e-175;  
 Matches 660; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY 1 atggagccctcaagggccggttagctttagcaagctggtcctgggtggtccgctgctc 60  
 DB 299 atggagccctcaagggccggttagctttagcaagctggtcctgggtggtccgctgctc 358  
 QY 61 ctgcctctggcgtggaagggctctgagctacccagataagcccaatgccaag 120  
 DB 359 ctgcctctggcgtggaagggctctgagctacccagataagcccaatgccaag 418  
 QY 121 agcgtgcaaaattgtcaaaagtgcctttattgtaaaagcaacagagactaagtct 180  
 DB 419 agcgtgcaaaattgtcaaaagtgcctttattgtaaaagcaacagagactaagtct 478  
 QY 181 catgccccttgcctgctgataagaggacacatctctgggtggtgatctcaagaact 240  
 DB 479 catgccccttgcctgctgataagaggacacatctctgggtggtgatctcaagaact 538  
 QY 241 tctctggagagaccctggtcacaacttcaacagcacataccactgltcaltagaact 300  
 DB 539 tctctggagagaccctggtcacaacttcaacagcacataccactgltcaltagaact 598  
 QY 301 caagcaaaacccctcaaaagtgtactgtgccaacacctctcggtgttactaagctcag 360  
 DB 599 caagcaaaacccctcaaaagtgtactgtgccaacacctctcggtgttactaagctcag 658  
 QY 361 acctatatactgcacaacatgcaactgtcctggaagaaatgaacctggaatactac 420  
 DB 659 acctatatactgcacaacatgcaactgtcctggaagaaatgaacctggaatactac 718  
 QY 421 acctatatactgcacaacatgcaactgtcctggaagaaatgaacctggaatactac 480  
 DB 719 acctatatactgcacaacatgcaactgtcctggaagaaatgaacctggaatactac 778  
 QY 481 gaccagaatgtgtctggaatgatactgtgtacccgtgagtgccaaagtcttttgcag 540  
 DB 779 gaccagaatgtgtctggaatgatactgtgtacccgtgagtgccaaagtcttttgcag 837  
 QY 541 tgtgttgcgtatgatttccatgatacaagtgtatgcgaagggcgttcttcaag 600  
 DB 838 tgtgttgcgtatgatttccatgatacaagtgtatgcgaagggcgttcttcaag 896  
 QY 601 ctatgttcttcggaattcttggagccacactctatccgtctcaatctgcttggag 660  
 DB 897 ctatgttcttcggaattcttggagccacactctatccgtctcaatctgcttggag 955  
 QY 661 acc 663  
 DB 956 acc 958

RESULT 14  
 BG432965 744 bp mRNA EST 14-MAR-2001  
 LOCUS 602495788F1 NIH\_MGC\_75 Homo sapiens cDNA clone IMAGE:4609773 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG432965  
 VERSION BG432965.1 GI:13339471  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 744)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: rgs@nih.gov  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 plate: LNCM353 row: 0 column: 22  
 High quality sequence stop: 744.  
 Location/Qualifiers  
 1. 744

## FEATURES

source  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4609773"  
 /clone\_lib="NIH\_MGC\_75"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /note="Organ: Kidney; Vector: pDNR-D1B (Clontech); Site: 1:  
 SfiI (ggcgcctggcc); Site 2: SfiI (ggcattatgccc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-ATTCTAGAGCGCGCGCGCGCATG-dt(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.65  
 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH\_MGC library."  
 BASE COUNT 181 a 199 c 177 g 187 t  
 ORIGIN

Query Match 91.0%; Score 625.4; DB 11; Length 744;  
 Best Local Similarity 99.4%; Pred. No. 1.4e-174;  
 Matches 659; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

QY 1 atggagccctcaagggccggttagctttagcaagctggtcctgggtggtccgctgctc 60  
 DB 10 atggagccctcaagggccggttagctttagcaagctggtcctgggtggtccgctgctc 69  
 QY 61 ctgcctctggcgtggaagggctctgagctacccagataagcccaatgccaag 120  
 DB 70 ctgcctctggcgtggaagggctctgagctacccagataagcccaatgccaag 128  
 QY 121 agcgtgcaaaattgtcaaaagtgcctttattgtaaaagcaacagagactaagtct 180  
 DB 129 agcgtgcaaaattgtcaaaagtgcctttattgtaaaagcaacagagactaagtct 188  
 QY 181 catgccccttgcctgctgataagaggacacatcttgggtgtgatactcagaactgt 240  
 DB 189 catgccccttgcctgctgataagaggacacatcttgggtgtgatactcagaactgt 248  
 QY 241 tctctggagagaccctggtcacaacttcaacagcacataccactgltcaltagaact 300  
 DB 249 tctctggagagaccctggtcacaacttcaacagcacataccactgltcaltagaact 308  
 QY 301 caagcaaaacccctcaaaagtgtactgtgccaacacctctcggtgttactaagctcag 360  
 DB 309 caagcaaaacccctcaaaagtgtactgtgccaacacctctcggtgttactaagctcag 368  
 QY 361 actctgatactgcacaacatgcaactgtctctggaagaaatgaactgcttgaatact 420  
 DB 369 actctgatactgcacaacatgcaactgtctctggaagaaatgaactgcttgaatact 428  
 QY 421 acctcttatagaacaacaaatctgtcaaggggcaaaagaaacctttgcaataacag 480  
 DB 429 acctcttatagaacaacaaatctgtcaaggggcaaaagaaacctttgcaataacag 488  
 QY 481 gaccagaatgtgtcctggaatgatactgtgttactatggtlccaggtcttttgcag 540  
 DB 489 gaccagaatgtgtcctggaatgatactgtgttactatggtlccaggtcttttgcag 548  
 QY 541 tgtgttgcgtatgatttccatgatacaagtgtatgcgaagggcgttcttcaag 600

QY 507 atcttgtaactgacatggtccaggtcttcttgacgtgtgttgcgtgacatggttccatg 566  
|||||  
Db 481 ATCTGTGTACCTGATGTCATGAGTCCAGGCTTTTGTGCACTGTGTGTGTCATGATGTTTCCATGG 540  
QY 567 atacaagtatgtagcgaagggctgtcttcacactctatgtcttcggagatctggagc 626  
|||||  
Db 541 ATACAGTATGATGCGCAGGCGCTGTCTCAGTCTTATGTCTTCGGGATTTCTGGAGC 600  
QY 627 caccactatcgcgtctccattctgtcttgagc 660  
|||||  
Db 601 CACCACCTATCCGCTCCATTCATCTGTGGGCG 634

RESULT 12  
AL560710 799 bp mRNA EST 16-FEB-2001  
LOCUS AL560710 LTI\_NFL010\_BC2 Homo sapiens cDNA clone CS0D1004YB05 3  
DEFINITION prime, mRNA sequence.  
ACCESSION AL560710 GI:12907431  
VERSION AL560710.1 GI:12907431  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Chordata: Craniata: Vertebrata: Euteleostomi:  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo:  
1 (bases 1 to 799)  
REFERENCE 1, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers

FEATURES  
source  
1. 799  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0D1004YB05"  
/clone\_1ib="LTI\_NFL010\_BC2"  
/sex="male"  
/tissue-type="B cells from Burkitt lymphoma"  
/note="Vector: pCMVSPORT 6; Site: 1: NotI; 1st strand cDNA.  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed by  
Life Technologies. Contact : Feng Liang Life Technologies,  
a division of Invitrogen 9800 Medical Center Drive  
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
Email : fliang@lifestech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 180 a 186 c 225 g 187 t 21 others

ORIGIN

Query Match 91.98; Score 631.4; DB 10; Length 799;  
Best Local Similarity 94.98; Pred. 2.5e-176;  
Matches 652; Conservative 18; Mismatches 15; Indels 2; Gaps 2;

QY 1 atggcgctacagccgggtaagctctacagccgtgctccgggtgctgcgcctgctc 60  
|||||  
Db 775 ATGGGCGCTCAGGCGCGGACT-TTACGACCTGTGCTCCCT-KGCTGCGCGCTGCCCC 718

QY 61 ctgcctctggcgtgtaaaaggtcttcgctacaccagatgacacccaatgtccagag 120  
|||||  
Db 717 CTGCTCTGGGCGTGAAGAGRCCTGCGCGMACCGAGATATGACACCAATGTCACAGG 658

QY 121 agcgtgcaaaattgtcaaaagtgcttattgttaaaagcagacagagagctaatgctg 180  
|||||  
Db 657 AGCGARMAAATTTATMAAAGTGRMTTATRTAAAGACAGACAGAGCTAAATGCTG 598

QY 181 catgccgttgctgcctgaatcagaagacacatcttggtgctgacatcccaactgt 240

Db 597 CATGCCCGTGTGCTGCTGATGATGAGAGGCGACCATCTTGGGGCTGATCTCCAGAACTGT 538  
|||||  
QY 241 tctctggaagacccttggtccaaacttcatcagacacatcactgtcatatagactg 300  
|||||  
Db 537 TCTCTGAGAGACCCCTGTCACAACTTTCATCAGGCATACVACGTGTCATATAGACCGG 478  
QY 301 caagaacacccctcaaaagtgacttggtgcaaacacttcgtgtgttactcagctcag 360  
|||||  
Db 477 CAAGCAAAACCCCTGAAAGTACTTGGCAACRCCCGCGGCTTCACTCAGCTCCAG 418  
QY 361 actctatctgcacaaatgtcaactgtcttcgtgaggaatgaagcttgaaactc 420  
|||||  
Db 417 ACTCTGATCTGCCACAAATCTCACTGCTCTGGAGGATTAATGCTGGAATACACTG 358  
QY 421 acccttataagacacccaatctcgtcaaggcgaagaaccttgcataaactgag 480  
|||||  
Db 357 ACTCTTATATAGACAMCCAAATCTGTCAAGGCAAAAGAMCCTTGCATTAACACTGG 298  
QY 481 gaccagaagaatgtccctggaagatgactgtgactgctgtgtccaggtctttgag 540  
|||||  
Db 297 GACCCAGAAATGTGCCCTGAGATGATCTGTGTCGATGATGTCAGGCTTTTTCAG 238  
QY 541 tgtgttgctgacgtgttccatgatacaagtgatgcgcaaggtctgtctcaactg 600  
|||||  
Db 237 TGTGTTTGCTGATGATGTTTCCATGATACAGTATATGCCGAGGCTCTTCACAG 178  
QY 601 ctatgtcttcggaattctggagacacacactatcgtctccatctgtcttgagcg 660  
|||||  
Db 177 CTATGTCTTCTGGGATCTGTGGAGCCACCTACCGTCCCATCTTGTGGGCG 118  
QY 661 acccagcgccgaagacagacttca 687  
|||||  
Db 117 ACCCAGCGCCGAAAGCAAGACTTCA 91

RESULT 13  
AL543210 969 bp mRNA EST 16-FEB-2001  
LOCUS AL543210 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0D1001YF08 5  
DEFINITION prime, mRNA sequence.  
ACCESSION AL543210 GI:12875688  
VERSION AL543210.1 GI:12875688  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Chordata: Craniata: Vertebrata: Euteleostomi:  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo:  
1 (bases 1 to 969)  
REFERENCE 1, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL  
COMMENT Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers

FEATURES  
source  
1. 969  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0D1001YF08"  
/clone\_1ib="LTI\_NFL006\_PL2"  
/tissue-type="Placenta"  
/note="Vector: pCMVSPORT 6; Site: 1: NotI; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed by  
Life Technologies. Contact : Feng Liang Life Technologies,  
a division of Invitrogen 9800 Medical Center Drive  
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
Email : fliang@lifestech.com URL :  
http://fulllength.invitrogen.com"



|||||  
Db 481 ACCGAGAAATGTCCTGAGATGATCTTGATACCTGATGTCACAGTCTTTTGACGT 540  
Qy 542 ggtcttgctgcagatggttccatgatacaagtgatgagccagggctcgtcttcacgc 601  
Db 541 GGTGTTGTCATGATGTTCCATGATACAAAGTATGCGCAGGGCTCGTTCACATGC 600  
Qy 602 ttatgtcttcggagttctggagagccacactctacgtcttcacatcttcgtctggagc-g 660  
Db 601 TTATGTTCTTCGCGATTTCTGGAGCCACACTCTATCCGCTCCATTTGCTTTGGCAG 660  
Qy 661 acccagcgccgaaagcccaagact 685  
Db 661 ACCCAGCGCCGAAAGCCCAAGACTT 685  
RESULT 9  
BF026265 952 bp mRNA EST 10-OCT-2000  
LOCUS 601672891F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:3955607 5',  
DEFINITION mRNA sequence.  
ACCESSION BF026265  
VERSION BF026265.1 GI:10733977  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 952)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DRP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov  
Plate: LNC850 row: f column: 24  
High quality sequence stop: 762.  
Location/Qualifiers  
1..952  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3955607"  
/clone\_id="NIH\_MGC\_20"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAAGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 257 a 227 c 237 g 230 t 1 others  
ORIGIN  
Query Match 92.8%; Score 637.6; DB 11; Length 952;  
Best Local Similarity 98.8%; Pred. No. 3.8e-178;  
Matches 664; Conservative 0; Mismatches 4; Indels 4; Gaps 2;  
Qy 16 ccggagtagcttaagaccctgtgctccctggagctgcgcgcctcctcgtctggagctg 75  
Db 2 CCGGTAATCTTACGACCCCTGCTGCTCCGCGCTCCGCTCCTCGCTCTGGCGCTG 61  
Qy 76 gaaagagctctgcgctacccgagatgacccaatgtccaggagcgtgcaaatltg 135  
Db 62 GAAAGGCGCTCGCGCTACCCGAGATATGACCAATGTCAGGAGCGAGCAAAATTG 121

Qy 136 tcaaaagtgccctttatgttaaaacagacagagacatagtgtatgccgttgctgc 195  
Db 122 TCAAAAGTGGCTTTTATGTTAAACGACACGAGACTTAATGTGATCCCGTGTCTGC 181  
Qy 196 ctgaatcagaagggacacacactctggggcctgcagctcccaagactctctctggagacc 255  
Db 182 CTGAATAGAGAGGCAACACTCTGGGGCTGGATCTCCAGACACTGTTCTTGAGAGACCT 241  
Qy 256 gttccaaacttcaagagacacacacacacacacacacacacacacacacacacac 315  
Db 242 GGTCCAAACTTTCATCAGGACATACACACTGTCATCATGAGACTGCACAAACCCCTTC 301  
Qy 316 aaagtgacttgcccaaacacactcgtgcttaccacacacacacacacacacacacac 375  
Db 302 AAAGGTACTTGGCCAAACACTTCCGCTGCTTACTGACTGACCTGACACTGTGATGCGCA 361  
Qy 376 caacatgcaacatgctcctggaggaatlaatgctcctggaaactacacacacacacacac 435  
Db 362 CAACATGTCAACGTCTCTGAGAGAAATTAATGCTGGAATATCATCACTTATATAGAC 421  
Qy 436 aaccacatctgtcaagggcaaaagaaaccttgcataacacatggggaaacagaaatgtg 495  
Db 422 AACCAATCTGTCAAGGGCAAAAGAAACCTTTGCAATACACTGGGGACCCCAAAATGTGT 481  
Qy 496 cctgagaatgactctgtgtacactgtagctgacaggtcttgcagtgctgtgtgt 555  
Db 482 CCGAGATGAGATCTGTGTACTGATGTCAGGCTTTCAGAGTG--TTGTGCTGAT 538  
Qy 556 ggttccatgatacaagtgatgacgagggcgtcttcacatgcttctcctcctcgg 615  
Db 539 GGTTCATGATGATACAAAGTATGAGCCGACGGGCTGTTCTCATGCTTATGTTCTGGG 598  
Qy 616 attctggagagcaacacactatcgtctccatctgtcttggggagccagcgcaaaa 675  
Db 599 ATTCTGGAGACCAACCACTATCCGTCTCATATCTGCTTTGGGCGA-CCAGCGCGGAAAA 657  
Qy 676 gcccaagacttca 687  
Db 658 GCCAAGACTTCA 669  
RESULT 10  
BE786718  
LOCUS BE786718 909 bp mRNA EST 20-OCT-2000  
DEFINITION 601475077F1 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:3878106 5',  
ACCESSION BE786718  
VERSION BE786718.1 GI:10207916  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 909)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: DCTD/DRP/Gazdar  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LMA9641 row: i column: 19  
High quality sequence stop: 708.  
Location/Qualifiers  
1..909  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
FEATURES  
source



Db 503 GACCAGAAATGTCCTGAGATGATCTTGTACCTGATGTCAGGCTTTTTCAG 562  
Qy 541 tgttttgcgtatgtttccatgatgataagatgataagcagagctgtctcactg 600  
Db 563 TGTGTTGTGCTGATGTTGCTCATGATACAGTGTATGCGCCAGGCTGTCTCAGT 622  
Qy 601 ctatgtctct-cgggattctggagcaccacactatccgtctcactcctgtctggc 659  
Db 623 CTTATGTTCTTCCGGGANTCTGGAGGACCAATCTAT-CGTCTCATTTCTGTTTGGC 681  
Qy 660 gaccagcgcgcaaaagcgaactc 686  
Db 682 GACCAGCGCCGAAAGCCAGACTTC 708

RESULT 6  
Bg826210 860 bp mRNA EST 22-MAY-2001  
LOCUS 602750603F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4903297 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg826210  
VERSION Bg826210.1 GI:14173797  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 860)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: rgs@bgl-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM801 row: f column: 02  
High quality sequence stop: 802.  
Location/Qualifiers  
1..860  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4903297"  
/clone\_lib="NIH\_MGC\_17"  
/tissue\_type="rhodomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pORF7; Site\_1: EcoRI;  
Site\_2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGACAGAG(9). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 206 a 230 c 206 g 218 t  
ORIGIN

Query Match 95.28; Score 654.2; DB 11; Length 860;  
Best Local Similarity 99.44; Pred. NO. 4.4e-183;  
Matches 667; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 10 caccgcccggtagctcttaacacccctggtgcccctggcgcgcgcctcctcgtc 69  
Db 2 CACGCGCCGGGTACTCTTACGACCCCTGTGCTGCGGCTGCGCCCTGCTCTGCTG 61  
Qy 70 ggcgtggaagagctctggtgcgtaccgagatatgacccaatgtccaggagcgtgca 129  
Db 62 GCGCTGGAAGAGGCTGTGGCGCTACCCGAGATATGACCAACCAATGTCCA-GGAGGTGCAA 120

Qy 130 aattgtcaaaagtggccttattgtlaaacaagacagagactaattgtcgtatgccgt 189  
Db 121 AATTGTCAAAAGTGGCTTTATTGTATAAAGACACAGACTAATGCTGATGCCGT 180  
Qy 190 tgcgtcctgaatcaagaaggacacactctgtggcctgagatcccaagtgtctctggag 249  
Db 181 TGTGCTGCTGAATCAAGAGGACCACTTGTGGGGCTGAGATCTCAAGACGTGTTCTGGAG 240  
Qy 250 gaecctgttccaaacttcaagacacataccactgttcatatcatagactgtgaagcaaac 309  
Db 241 GACCCGTGTCAAACTTTATCATGAGCATACACCTGCTCATATGAGACCTGCAAGCAAC 300  
Qy 310 cccctcaaaagtgaacttggcacaacacttcggtgcttactcaagctcagaactgta 369  
Db 301 CCCCTAAAGCTGACTTGGCCCAACACTTCCGTGCTTACTGACTCCAGACTGTGATA 360  
Qy 370 ctgcacaacatgttcaactgtctctgaggaattaaatgctggaatacactacactat 429  
Db 361 CTGCCACAAACATGTCAAGTCTCTGGAGGAAATTAATGCTGGAATATATCACTTTAT 420  
Qy 430 atagacaaccaaatctgttcaaggagcaaaagaacctttgcaatacaactgggagccagaa 489  
Db 421 ATAGACAAACCAATCTGTCAAGGGCAAAAGAACTTTGCAATTAACACTGGGACCCAGAA 480  
Qy 490 atgtgtccctgagaatgagatctgtgtacactgagtgagtgagcaggtctttgcaatgtgtgt 549  
Db 481 ATGTGTCTGTGAGATGAGATCTGTGTACTGTGATGGTCCAGGCTTTTGTGACTGTGTGTGT 540  
Qy 550 gctgtagttccatgtgatacaagtgatgacgacagagctgtctcactgtctatgttc 609  
Db 541 GCTGATGTTTCATGATGATCAAGGTGTATGCCAGGGCTCGTTCTCACTGTTATGTTTC 600  
Qy 610 ttggagatctctggagagcaccactctatccgtctcactgtcttggcgagccagagc 669  
Db 601 TTCGGATTTCTGGAGACCAACACTATTCCTCATTTCTGTTGGCGACCCAGCGC 660  
Qy 670 cgaagaagccaa 680  
Db 661 CGAAAGAGCCA 671

RESULT 7  
AL553473 874 bp mRNA EST 16-FEB-2001  
LOCUS AL553473 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0D1076YH14 5  
DEFINITION partial mRNA sequence.  
ACCESSION AL553473  
VERSION AL553473.1 GI:12893340  
KEYWORDS E.  
SOURCE hu.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 874)  
AUTHORS L. Gruber, C. J. Jessup, J. and Polayes, D.  
TITLE Full length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Cor'act: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers  
1..874  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0D1076YH14"  
/clone\_lib="LTI\_NFL006\_PL2"  
/tissue\_type="placenta"  
/note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6





QY 601 ctatgtcttcggaattctcggagagccaccactctacccgtctcattcttcttggcg 660  
 |||  
 Db 622 CTATGTTCTTCGGGATTCTGGGAGCCACACTATCCGTCTCATCTTGTCTTGGCG 681  
 |||  
 QY 661 acccagcgcgaaagcccaagacttca 687  
 |||  
 Db 682 ACCCAGCGCGGAAAAGCCAAAGACTTCA 708  
 |||

RESULT 3  
 AF077037 923 bp mRNA HTC 22-MAY-2001  
 LOCUS AF077037 Homo sapiens HSPC013 mRNA, complete cds.  
 DEFINITION AF077037  
 ACCESSION AF077037  
 VERSION AF077037.1 GI:4689121  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 923)  
 AUTHORS Zhang, Q.H., Ye, M., Wu, X.Y., Ren, S.X., Zhao, M., Zhao, C.J., Fu, G.,  
 Shen, Y., Fan, H.Y., Lu, G.Y., Zhong, M., Xu, X.R., Han, Z.G., Zhang, J.W.,  
 Tao, J., Huang, Q.H., Zhou, J., Hu, G.X., Gu, J., Chen, S.J. and Chen, Z.  
 Cloning and functional analysis of cDNAs with open reading frames  
 for 300 previously undefined genes expressed in CD34+ hematopoietic  
 stem/progenitor cells  
 JOURNAL Genome Res. 10 (10), 1546-1560 (2000)  
 MEDLINE 20499367  
 PUBMED 11042152  
 REFERENCE 2 (bases 1 to 923)  
 AUTHORS Zhou, J., Zhang, Q., Fu, G., Ye, M., Yu, Y., Shen, Y., Xu, S., He, K.,  
 Chen, S., Mao, M. and Chen, Z.  
 Human HSPC013 mRNA  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 923)  
 AUTHORS Zhou, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-JUL-1998) Shanghai Institute of Hematology, Shanghai  
 Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,  
 Shanghai 200025, P.R. China  
 FEATURES  
 SOURCE location/Qualifiers  
 1..923  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="CD34+ cell"  
 195..710  
 /codon\_start=1  
 /product="HSPC013"  
 /protein\_id="AAD27770.1"  
 /db\_xref="GI:4689122"  
 /translation="MLHARCCLNKQITLLGDLDNCSLEDPGRFRAHMYVITIDIA  
 NPLKGLANFRFTLOTLILPOHVNCPGGINAMNTTITIDNIOGKNCNNTG  
 DPENCENGSCVDPGPGILQCVADGPHGKCMQSGSFLMFGILGAVTLVSVSTL  
 WATQRRKRTS"  
 CDS  
 BASE COUNT 227 a 244 c 217 g 235 t  
 ORIGIN  
 Query Match 98.3%; Score 675; DB 12; Length 923;  
 Best Local Similarity 99.9%; Pred. No. 3.1e-189;  
 Matches 686; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 141 AGCGTCGCAAAATTTGTCAAAAGTGCGCTTTATGTAAGACGACGAGACTAATGCTG 200  
 |||  
 QY 181 catgcgcgttgcctgctgaatcagaagggcaccaatcttggggctggatctccaagact 240  
 |||  
 Db 201 CATGCCCGTTCGCTCATATGAGAAAGGACCACTTCTGGGCGATCTCCAGAACTGT 260  
 |||  
 QY 241 tctctggagagaccctggtccaaacttcatcagacatcagatcagatcagactg 300  
 |||  
 Db 261 TCTCTGGAGGACCTTGCTCAAACTTTTCATCGACGACATACATCATCATATGAGCTG 320  
 |||  
 QY 301 caagcaaacccctcaaaagtgacttgccaaacacttcgctgacttactcagctcag 360  
 |||  
 Db 321 CAAGCAAAACCCCTCAAAAGTGACTTGGCCAAACACTTCCTGCTTATCAGCTCAG 380  
 |||  
 QY 361 actctgatactgcacaaatcagatcagctgctcctggaggaataagcttgatactc 420  
 |||  
 Db 381 ACTCTGATACCTGCCAACATCTCAATCTCTCTGAGCAATTAATGCTTGGAATCTATC 440  
 |||  
 QY 421 acccttatatagacaaccaaatctgtcaaggcacaagaacacttgcataacacatg 480  
 |||  
 Db 441 ACCTCTTATATAGACACCAAAATCTGTACAGGCCAAAGAACCTTGCAATACACTGG 500  
 |||  
 QY 481 gaccagaaatgtgtctcgaagaatgatactgtgttaactcgtatgctccaggtcttgcag 540  
 |||  
 Db 501 GACCCAGAAATGTCTCTAGAAATGATCTGTGTACCTGATGTCACAGCTTTTGCAG 560  
 |||  
 QY 541 tgtgttgctgatagtttcattcagatagaatgaatgatacagatgacgacaggtcttccatg 600  
 |||  
 Db 561 TGTGTTTGCTGATGCTTCCATGATGATACACTGTATGCGCAGGCGCTTTCACAG 620  
 |||  
 QY 601 ctatgtcttcggaattctcggagagccaccactctacccgtctcattcttcttggcg 660  
 |||  
 Db 621 CTATGTTCTTCGGGATTCTGGGAGCCACACTATCCGTCTCATCTTGTCTTGGCG 680  
 |||  
 QY 661 acccagcgcgaaagcccaagacttca 687  
 |||  
 Db 681 ACCCAGCGCGGAAAAGCCAAAGACTTCA 707  
 |||

RESULT 4  
 AL569918/c 902 bp mRNA EST 16-FEB-2001  
 LOCUS AL569918 LTI\_NFL006.PL2 Homo sapiens cDNA clone CSDD1001.F08 3  
 DEFINITION prime, mRNA sequence.  
 ACCESSION AL569918  
 VERSION AL569918.1 GI:12925734  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 902)  
 AUTHORS Li, W.B., Gruber, C., Jesssee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
 FEATURES  
 SOURCE location/Qualifiers  
 1..902  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CSDD1001.F08"  
 /clone\_1lb="LTI\_NFL006.PL2"  
 /tissue\_type="placenta"  
 /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA  
 was primed with a NotI-Oligo(dT) primer. Five prime end  
 clones, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library was normalized. Library was constructed by  
 Life Technologies. Contact : Feng Liang Life Technologies,  
 a division of Invitrogen 9800 Medical Center Drive



Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.5kb. Library constructed by Life  
Technologies. Note: this is a NCI-CGAP Library."

BASE COUNT 235 a 248 c 211 g 235 t

ORIGIN

Query Match 100.0%; Score 687; DB 11; Length 929;  
Best Local Similarity 100.0%; Pred. No. 8,5e-193;  
Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atggcgctcaccgcccgggtagtcttaacgacctggtgctggtcgccgctgtc 60
DB 5 atggcgctcaccgcccgggtagtcttaacgacctggtgctggtcgccgctgtc 64
QY 61 ctgctgctggtggtgaaaggctctgctgctgctgctgctgctgctgctgctg 120
DB 65 ctgctgctggtggtgaaaggctctgctgctgctgctgctgctgctgctgctg 124
QY 121 agcgtgcaaaatttgcaaaagtggccttattgttaaaacgacagagagctatgctg 180
DB 125 accgtgcaaaatttgcaaaagtggccttattgttaaaacgacagagagctatgctg 184
QY 181 catgcccgttgcctgctgaatcagaaggacacatcttggggtggtgctgctcagaactgt 240
DB 185 catgcccgttgcctgctgaatcagaaggacacatcttggggtggtgctgctcagaactgt 244
QY 241 tctctggaggaacctggtgcaaaacttcatcagagacatacactgctatcattgactgt 300
DB 245 tctctggaggaacctggtgcaaaacttcatcagagacatacactgctatcattgactgt 304
QY 301 caagcaaaacccctcaaaagtgtactgtgcaacacactcgtggtcttaccacacacac 360
DB 305 caagcaaaacccctcaaaagtgtactgtgcaacacactcgtggtcttaccacacacac 364
QY 361 actctgatactgcccacacatgtcaactgtctcgtgaggaatlaatgctggaatactatc 420
DB 365 actctgatactgcccacacatgtcaactgtctcgtgaggaatlaatgctggaatactatc 424
QY 421 acccttataagacaacaaatctgtcaaggacaaagaaaccttggaaataaactgag 480
DB 425 acccttataagacaacaaatctgtcaaggacaaagaaaccttggaaataaactgag 484
QY 481 gaaccaagaatgtgtcctgagaatgagatctgtgtacatgagtcgaagtccttgcag 540
DB 485 gaaccaagaatgtgtcctgagaatgagatctgtgtacatgagtcgaagtccttgcag 544
QY 541 tgtgttggctgagtggttccatgagataaagtgtatgctgcccagggctgttctactgt 600
DB 545 tgtgttggctgagtggttccatgagataaagtgtatgctgcccagggctgttctactgt 604
QY 601 ctatgtctctcgtgagatctggaagcaccactatcgtctcattcgtctgctgagcg 660
DB 605 ctatgtctctcgtgagatctggaagcaccactatcgtctcattcgtctgctgagcg 664
QY 661 acccaagcgccgaaaaagcaacttca 687
DB 665 acccaagcgccgaaaaagcaacttca 691

```

RESULT 2

AL560711 938 bp mRNA EST 16-FEB-2001

LOCUS AL560711 LTI\_NF01010-BC2 Homo sapiens cDNA clone CSD0L004YB05 5

DEFINITION prime, mRNA sequence.

ACCESSION AL560711

VERSION AL560711.1 GI:12907433

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 938)

AUTHORS Li, W. B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: [seqlife@genoscope.cns.fr](mailto:seqlife@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).

FEATURES

source

1..938

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CS0DL004YB05"

/clone\_1ib="LTI\_NF01010-BC2"

/sex="male"

/tissue\_type="B cells from Burkitt lymphoma"

/note="Vector: pCMVSPORT 6; Site-1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : [liang@life.com](mailto:liang@life.com) URL : <http://fulllength.invitrogen.com>"

BASE COUNT 238 a 244 c 217 g 234 t 5 others

ORIGIN

Query Match 99.8%; Score 685.4; DB 10; Length 938;  
Best Local Similarity 99.4%; Pred. No. 2,5e-192;  
Matches 683; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atggcgctcaccgcccgggtagtcttaacgacctggtgctggtcgccgctgtc 60
DB 22 atggcgctcaccgcccgggtagtcttaacgacctggtgctggtcgccgctgtc 81
QY 61 ctgctgctggtggtgaaaggctctgctgctgctgctgctgctgctgctgctg 120
DB 82 ctgctgctggtggtgaaaggctctgctgctgctgctgctgctgctgctgctg 141
QY 121 agcgtgcaaaatttgcaaaagtggccttattgttaaaacgacagagagctatgctg 180
DB 142 agcgtgcaaaatttgcaaaagtggccttattgttaaaacgacagagagctatgctg 201
QY 181 catgcccgttgcctgctgaatcagaaggacacatcttggggtggtgctgctcagaactgt 240
DB 202 catgcccgttgcctgctgaatcagaaggacacatcttggggtggtgctgctcagaactgt 261
QY 241 tctctggaggaacctggtgcaaaacttcatcagagacatacactgctatcattgactgt 300
DB 262 tctctggaggaacctggtgcaaaacttcatcagagacatacactgctatcattgactgt 321
QY 301 caagcaaaacccctcaaaagtgtactgtgcaacacactcgtggtcttaccacacacac 360
DB 322 caagcaaaacccctcaaaagtgtactgtgcaacacactcgtggtcttaccacacacac 381
QY 361 actctgatactgcccacacatgtcaactgtctcgtgaggaatlaatgctggaatactatc 420
DB 382 actctgatactgcccacacatgtcaactgtctcgtgaggaatlaatgctggaatactatc 441
QY 421 acccttataagacaacaaatctgtcaaggacaaagaaaccttggaaataaactgag 480
DB 442 acccttataagacaacaaatctgtcaaggacaaagaaaccttggaaataaactgag 501
QY 481 gaaccaagaatgtgtcctgagaatgagatctgtgtacatgagtcgaagtccttgcag 540
DB 502 gaaccaagaatgtgtcctgagaatgagatctgtgtacatgagtcgaagtccttgcag 561
QY 541 tgtgttggctgagtggttccatgagataaagtgtatgctgcccagggctgttctactgt 600
DB 562 tgtgttggctgagtggttccatgagataaagtgtatgctgcccagggctgttctactgt 621

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2002, 23:32:28 ; Search time 1499.8 Seconds  
(without alignments)  
4922.223 Million cell updates/sec

Title: US-09-726-348-1\_COPY\_5\_691

Perfect score: 687  
Sequence: 1 atggcgccctacgcgcgcg99.....gccgaagaagccaagacttca 687

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: em\_estfun: \*  
2: em\_estlin: \*  
3: em\_estlin: \*  
4: em\_estlin: \*  
5: em\_estlin: \*  
6: em\_estlin: \*  
7: em\_estlin: \*  
8: em\_estlin: \*  
9: em\_estlin: \*  
10: gb\_estlin: \*  
11: gb\_estlin: \*  
12: gb\_hlc: \*  
13: gb\_gss: \*  
14: em\_gss\_fun: \*  
15: em\_gss\_fun: \*  
16: em\_gss\_inv: \*  
17: em\_gss\_pln: \*  
18: em\_gss\_pro: \*  
19: em\_gss\_pro: \*  
20: em\_gss\_vrt: \*  
21: em\_gss\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	687	100.0	929	11	BG674532 602620574
2	685.4	99.8	938	10	AL560711
3	675	98.3	923	12	AF077037 Homo sapi
4	664.8	96.8	902	10	AL569918
5	658.2	95.8	834	11	BG506779
6	654.2	95.2	860	11	BG826210
7	650	94.6	874	10	AL553473
8	649	94.5	881	11	BG471234
9	637.6	92.5	952	11	BF026265
10	635.8	92.1	909	10	BE786718
11	632.4	92.1	634	10	BE409428
12	631.4	91.9	799	10	AL560710

13	627	91.3	969	10	AL543210
14	625.4	91.0	744	11	BG432965
15	625.2	91.0	980	10	AL527415
16	621.6	90.5	810	11	BL255439
17	618	90.0	724	10	BE298514
18	615	89.5	805	11	BG818135
19	612.8	89.2	909	11	BF026624
20	601.6	87.6	716	10	BE728887
21	595.8	86.7	898	11	BE902516
22	590.4	85.9	941	10	AL582873
23	590.4	85.9	961	10	AL582922
24	588	85.6	814	11	BG196385
25	587.8	85.6	923	10	BE781883
26	586	85.3	624	10	BE409758
27	583.6	84.9	903	11	BG036611
28	583.4	84.9	887	10	AL543302
29	581.6	84.7	800	11	BG213523
30	577.6	84.1	868	10	AL553441
31	577.2	84.0	1107	11	BG425473
32	576.8	84.0	847	10	AL522977
33	576	83.8	736	10	AL522976
34	564.2	82.1	983	11	BF568872
35	562.2	81.8	845	11	BF575660
36	543.4	79.1	770	10	AL562888
37	527.6	76.8	842	11	BG207774
38	523.6	76.2	901	11	BG761125
39	516.6	75.2	926	11	BG756959
40	508.2	74.0	595	10	AV714282
41	505.6	73.6	914	11	BE471266
42	499	72.6	946	10	BE783839
43	498.8	72.6	893	11	BF685464
44	493	71.8	747	11	BG675772
45	492	71.6	553	11	BG329351

#### ALIGNMENTS

RESULT 1  
BG674532 929 bp mRNA EST 01-MAY-2001  
LOCUS 602620574F1 NCI\_CGAP\_Skn3 Homo sapiens cDNA clone IMAGE:4745826 5',  
DEFINITION mRNA sequence.  
ACCESSION BG674532  
VERSION BG674532.1 GI:13905928  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
1 (Bases 1 to 929)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: James Cleaver, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA  
Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLNL0593 row: d column: 19  
High quality sequence stop: 844.  
Location/Qualifiers  
1. 929  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4745826"  
/clone\_1b="NCI CGAP\_Skn3"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin; Vector: pCMV-Sport6; Site: 1; NCI;"

#### FEATURES

source



Thu Mar 28 09:21:34 2002

us-09-726-348-2\_1.rpt

---

Db 1284 VCPGFGVARG 1294

RESULT 14  
ID 097458 PRELIMINARY; PRT; 2704 AA.

AC 097458;  
DT 01-MAY-1999 (TREMblrel. 10, Created)  
DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE EG:163A10.2 PROTEIN.  
GN N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Borkova D., Minana B., Kafatos F.C.;  
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Benos P.;  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL035395; CAB37611.1; -;  
DR EMBL; AL035436; CAB37611.1; JOINED.  
DR EMBL; AL035436; CAB37610.1; -;  
DR HSSP; P00740; 1EDM.  
DR FLYbase; FBgn004647; N.  
DR Interpro: IPR002110; ANK.  
DR Interpro: IPR000152; Asx\_hydroxyl.  
DR Interpro: IPR000561; EGF-like.  
DR Interpro: IPR000742; EGF-2.  
DR Interpro: IPR001881; EGF-Ca.  
DR Interpro: IPR001438; EGF-II.  
DR Interpro: IPR002049; Laminin\_EGF.  
DR Interpro: IPR000800; Notch.  
DR Pfam; PF000023; ank; 6.  
DR Pfam; PF00008; EGF; 36.  
DR Pfam; PF00066; notch; 3.  
DR PRINTS; PR00010; EGFBL0D.  
DR PRINTS; PR00011; EGFAMININ.  
DR PRINTS; PR01452; NOTCH.  
DR SMART; SM00248; ANK; 4.  
DR SMART; SM00179; EGF\_CA; 23.  
DR SMART; SM00004; EGF\_like; 13.  
DR PROSITE; PS50088; ANK\_REPEAT; 5.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_34.  
DR PROSITE; PS01186; EGF\_2; 28.  
DR PROSITE; PS01187; EGF\_CA; 21.  
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
KW SEQUENCE 2704 AA; 288894 MW; 554298A25C9EBA4E CRC64;

Query Match 8.4%; Score 105; DB 5; Length 2704;  
Best Local Similarity 24.6%; Pred. No. 0.052;  
Matches 47; Conservative 17; Mismatches 59; Indels 68; Gaps 10;

QY 27 RALALPEICTOCGSGVQNLNRYAF-----YCK-----TTRRLMLHAR 63  
DB 1147 KGLISRLQDNNCTCKDYGNHVCYSGYAGSYCKEIDECQSPQCGNGTCRDLGAYE 1206  
QY 64 CCLNKGRTLLGLDNLCSLEDPGPNFHOAHHTVILIDLANPLK--GDLANTFRGFTLOQT 121  
DB 1207 CCGCRQ-----GFGGNGCNELN-----IDDCAPNCGQNGTCHDRVAMFS----- 1244

QY 122 LILPQHVNCPGINAMWNTITSYDNOICGGKNLNNTGPEMCPENGSCVPDPGLGL 181  
DB 1245 -----CSCPPG-----TWGL---ICEINKDDCK---PGACHNNGSCT-DRVGGFEC 1283  
QY 182 VCADGFFHGYKC 192  
DB 1284 VCPGFGVARG 1294

RESULT 15  
ID 09WU08 PRELIMINARY; PRT; 2872 AA.

AC 09WU08;  
DT 01-NOV-1999 (TREMblrel. 12, Created)  
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE FIBRILIN-1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99032689; PubMed=9815129;  
RA Kanwar Y.S., Ola K., Yang Q., Kumar A., Wada J., Kashihara N.,  
RA Peterson D.R.;  
RT "Isolation of rat fibrillin-1 cDNA and its relevance in metabolic  
development.";  
RL Am. J. Physiol. 275:F710-F723(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kanwar Y.S.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP135059; AAD34438.1; -;  
DR HSSP; P35535; IAPJ.  
DR Interpro: IPR000152; Asx\_hydroxyl.  
DR Interpro: IPR000561; EGF-like.  
DR Interpro: IPR001881; EGF-Ca.  
DR Interpro: IPR002212; TB.  
DR Interpro: IPR000822; znf-C2H2.  
DR Pfam; PF00008; EGF; 46.  
DR Pfam; PF00683; TB; 9.  
DR SMART; SM00179; EGF\_CA; 41.  
DR SMART; SM00004; EGF\_like; 5.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 42.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_2.  
DR PROSITE; PS01186; EGF\_2; 38.  
DR PROSITE; PS01187; EGF\_CA; 41.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_1.  
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
KW SEQUENCE 2872 AA; 312069 MW; 0C4F7F3B87A80280 CRC64;

Query Match 8.4%; Score 104; DB 11; Length 2872;  
Best Local Similarity 26.0%; Pred. No. 0.07;  
Matches 40; Conservative 11; Mismatches 57; Indels 46; Gaps 8;

QY 35 CTCPGSGVQNLNRYAFCTTRRLMLHARCLNKGRTLLGLDNLCSLEDPGPNFHOAH 93  
DB 2449 CNOAP-----KPCNFCNKTE---GSYQSCPCGYILQEDRSCKRDKDECATKQHNQ 2498  
QY 94 TTVIILQANPLKGDLANTFRGFTLOQTLLPQHVNCPGINAMWNTITSYDNOICGGK 153  
DB 2499 FLCY-----NTIGFT-----CKCPRGFTQHN--TACIDNNECTSBI 2533  
QY 154 NLNNTGDPPEMCPENGSCVPDPGLQVCADGF 187  
DB 2534 NLGSGSK---VCNT-----DGSFTCEQGRGF 2557

Search completed: March 28, 2002, 09:13:31  
Job time: 885 sec

DR HSP, P35555, 1APJ.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000257; Chitin\_binding.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR000212; TB.  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF00006; EGF\_64.  
 DR Pfam: PF00683; TB; 12.  
 DR SMART: SM00494; Chcd2; 2.  
 DR SMART: SM00179; EGF\_Ca; 60.  
 DR SMART: SM00001; EGF-like; 4.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 61.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01186; EGF\_2; 50.  
 DR PROSITE: PS01187; EGF\_CA; 61.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_1.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Repeat.  
 KW SEQUENCE 3857 AA; 418303 MW; 5BC0618BC527E04C CRC64;

Query Match 8.5%; Score 105.5; DB 11; Length 3857;  
 Best Local Similarity 21.7%; Pred. No. 0.067;  
 Matches 47; Conservative 19; Mismatches 72; Indels 79; Gaps 11;

OY 1 MAPHSGSLTLPWAAALLALVERLALPEICT--QC--PGSVONLSKVA----- 49  
 DB 3375 LCPHGRGEMTN-----GADIDCKVLDHCRNGECVNDGRSHYCKICTKTYTDPIT 3424  
 OY 50 -----FYCKTRRLMLHARCLNCKGTLLGLDQNC-SLEDPGPNFH 90  
 DB 3425 GTACVDLMECNAPKPCFKNTKNE---GSYQSCPRGYTLQEDGRCKDLDECAKQH 3480  
 OY 91 QAHTTVIIDLANPLKGLDANFEGTQLQTLILPQHVNCPEGINAMTTSYDNIQ 150  
 DB 3481 NCQFLCV-----NTIGGT-----CKCPGPTQHH--TACIDNNECT 3515  
 OY 151 GQKNCNNTGDEPCENGSCVPDGLQCVADGF 187  
 DB 3516 SDINLGSKG--VCQNT-----PGSFTECGORGF 3542

RESULT 13  
 O9W4T8 PRELIMINARY; PRT: 2703 AA.  
 AC O9W4T8; 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE N GENE PRODUCT.  
 GN N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolintinas S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,  
 RA Burlis K.C., Busam D.A., Butler H., Caden E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glotzer A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Klupp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195 (2000).  
 DR EMBL: AE003426; AAF45848.2; -.  
 DR HSP, P00740; 11XA.  
 DR Flybase: FBgn004647; N.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_1.  
 DR InterPro: IPR002049; Lamln1\_EGF.  
 DR InterPro: IPR000800; Notch.  
 DR Pfam: PF00023; ank; 6.  
 DR Pfam: PF00008; EGF; 36.  
 DR Pfam: PF00066; notch; 3.  
 DR PRINTS: PRO0010; EGFBLD.  
 DR PRINTS: PRO0011; EGFAMININ.  
 DR PRINTS: PRO1452; NOTCH.  
 DR SMART: SM00248; ANK; 4.  
 DR SMART: SM00179; EGF\_Ca; 23.  
 DR SMART: SM00001; EGF-like; 13.  
 DR PROSITE: PS00088; ANK\_REPEAT; 5.  
 DR PROSITE: PS0297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_34.  
 DR PROSITE: PS01186; EGF\_2; 28.  
 DR PROSITE: PS01187; EGF\_CA; 21.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SO SEQUENCE 2703 AA; 288851 MW; 0EAE23F426FECDB CRC64;

Query Match 8.4%; Score 105; DB 5; Length 2703;  
 Best Local Similarity 24.6%; Pred. No. 0.052;  
 Matches 47; Conservative 17; Mismatches 59; Indels 68; Gaps 10;

OY 27 RALALPEICQCGSVONLSKVAF-----YCK-----TTREMLHAR 63  
 DB 1147 KGLSLKRLCNNGCKKDYGNHVCYQAGYCYQAEIDECOSQPCQNGTGRDLGAE 1206  
 OY 64 CCLNCKGTLLGLDQNCSTEDPGPNHQAHTTVIIDLANPLK--GDLANTREGFTLOLT 121  
 DB 1207 CQCRQ-----GFGQNCCLN-----IDCCAPNCPQNGSTCHDRVANTS----- 1244  
 OY 122 LILPQHVNCPEGINAMTTSYDNIQCGQKYLKNNTGDEPCENGSCVPDGLQCV 181  
 DB 1245 -----CSCPFG-----TMGI-----ICETNKDDCK-----PGACHNNGSCI-DRYGGFEC 1283  
 OY 182 VCADGFGHYKGC 192

RN EMBL; U95299; AAC32288.1; -.  
 RP SEQUENCE FROM N.A.  
 RA Li L., Huang G., Banta A., Deng Y., Chen L., Rowen L.,  
 RA Hood L.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U95299; AAC32288.1; -.  
 DR HSP; P08709; 1BP9.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_11.  
 DR InterPro: IPR000800; Notch.  
 DR Pfam: PF000023; ank; 6.  
 DR Pfam: PF000066; EGF; 27.  
 DR Pfam: PF000066; notch; 2.  
 DR PRINTS; PR01415; ANKYRIN.  
 DR PRINTS; PR00010; EGFBLD.  
 DR PRINTS; PR01452; NOTCH.  
 DR SMART; SM00248; ANK; 5.  
 DR SMART; SM00179; EGF\_CA; 11.  
 DR SMART; SM00004; EGF\_1like; 15.  
 DR PROSITE; PS50088; ANK\_REPEAT; 5.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 11.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_28.  
 DR PROSITE; PS01186; EGF\_2; 21.  
 DR PROSITE; PS01187; EGF\_CA; 9.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 KW SEQUENCE 1999 AA; 209134 MW; A9787027EDA4E92A CRC64;

Query Match 8.6%; Score 106.5; DB 4; Length 1999;  
 Best Local Similarity 20.0%; Pred. No. 0.026;  
 Matches 47; Conservative 21; Mismatches 68; Indels 99; Gaps 12;

QY 2 APHGP-----GSLTLVPMALALALGERALALPEICTQ--CPGSVQNLKSKVA 49  
 DB 823 SPQGRPLCPPTGYTGSGSCQTLM-----DLCAQKPCPRN----- 855  
 QY 50 FYCKTTRELMHARCCNLKNGKTIIGLDLONCS-----LEDPGNPH 90  
 DB 856 SHCLQTPG-SFHCLCLOGMTGPLICNLPLSSCQKALSGIDVSSILCHNGGLCVDSGPXY- 913  
 QY 91 QAHVTIIDLQANPLKGLDANLTFRGFTQTLILPQHN-----CPGGINAMNTITSYI 144  
 DB 914 -----FCHCPRGF---QGSICQDHVNPCESRPCONGATCMAQPSGYL 952  
 QY 145 -----DNQICGQKNLNNNTGDEMCPEKSCVDPGRLQCCVADGFHGYKC 192  
 DB 953 CQCAPGYDGONCSKELDACOS---QPCNHGTCPTPK-PGGFHCACPPGFVGLRC 1002

RESULT 11  
 ID 000306 PRELIMINARY; PRT; 2003 AA.  
 AC 000306;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE NCBI\_TaxID=9606;  
 GN NCORCH4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BONE MARROW, AND HEART;  
 RA Li L., Huang G., Banta A., Yu D., Rowen L., Hood L.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U95299; AAC32288.1; -.  
 DR HSP; P08709; 1BP9.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_11.  
 DR InterPro: IPR000800; Notch.  
 DR Pfam: PF000023; ank; 6.  
 DR Pfam: PF000066; EGF; 27.  
 DR Pfam: PF000066; notch; 2.  
 DR PRINTS; PR01415; ANKYRIN.  
 DR PRINTS; PR01452; NOTCH.  
 DR SMART; SM00248; ANK; 5.  
 DR SMART; SM00179; EGF\_CA; 11.  
 DR SMART; SM00004; EGF\_1like; 15.  
 DR PROSITE; PS50088; ANK\_REPEAT; 5.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 11.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_28.  
 DR PROSITE; PS01186; EGF\_2; 21.  
 DR PROSITE; PS01187; EGF\_CA; 9.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 KW SEQUENCE 2003 AA; 209620 MW; BBD10283443A0C14 CRC64;

Query Match 8.6%; Score 106.5; DB 4; Length 2003;  
 Best Local Similarity 20.0%; Pred. No. 0.026;  
 Matches 47; Conservative 21; Mismatches 68; Indels 99; Gaps 12;

QY 2 APHGP-----GSLTLVPMALALALGERALALPEICTQ--CPGSVQNLKSKVA 49  
 DB 824 SPQGRPLCPPTGYTGSGSCQTLM-----DLCAQKPCPRN----- 856  
 QY 50 FYCKTTRELMHARCCNLKNGKTIIGLDLONCS-----LEDPGNPH 90  
 DB 857 SHCLQTPG-SFHCLCLOGMTGPLICNLPLSSCQKALSGIDVSSILCHNGGLCVDSGPXY- 914  
 QY 91 QAHVTIIDLQANPLKGLDANLTFRGFTQTLILPQHN-----CPGGINAMNTITSYI 144  
 DB 915 -----FCHCPRGF---QGSICQDHVNPCESRPCONGATCMAQPSGYL 953  
 QY 145 -----DNQICGQKNLNNNTGDEMCPEKSCVDPGRLQCCVADGFHGYKC 192  
 DB 954 CQCAPGYDGONCSKELDACOS---QPCNHGTCPTPK-PGGFHCACPPGFVGLRC 1003

RESULT 12  
 ID 088840 PRELIMINARY; PRT; 3857 AA.  
 AC 088840;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE MUTANT FIBRILLIN-1.  
 GN FBN1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B10.D2;  
 RX MEDLINE=98069008; PubMed=9405934;  
 RA Bona C.A., Murali C., Casares S., Kasturi K., Nishimura H., Honjo T.,  
 RA Matsuda F.;  
 RT "Structure of the mutant fibrillin-1 gene in the tight skin (TSK)  
 mouse.";  
 RL DNA Res. 4:267-271(1997).  
 DR EMBL; AF007248; AAC62317.1; -.

Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalili M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merklow G., Mlshina N.V., Mobarry C., Morris J., Mostreli A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J.D., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Sytkas R., Tector C., Turner R., Venler E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,"  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
CC - SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A (LDLRA) DOMAIN.  
EMBL: AF003615: AAF52472.1; -  
DR HSP: P00740: 1EDM.  
DR FlyBase: FBgn0031879; SPL070.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000859; CUB.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR000742; EGF-2.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR001438; EGF\_T1.  
DR InterPro: IPR000421; FA58\_C.  
DR InterPro: IPR001092; HLH\_dim.  
DR InterPro: IPR003410; HYR.  
DR InterPro: IPR001791; Laminin\_G.  
DR InterPro: IPR0002172; LDL\_recept\_A.  
DR InterPro: IPR000436; Sush1\_SCR\_CCP.  
DR InterPro: IPR001368; TNFR\_C6.  
DR Pfam: PF00431; CUB; 3.  
DR Pfam: PF00008; EGF; 16.  
DR Pfam: PF00754; F5\_F8\_type\_C; 1.  
DR Pfam: PF02494; HYR; 3.  
DR Pfam: PF00057; Idl\_recept\_a; 1.  
DR Pfam: PF00084; sush1; 7.  
DR PRINTS: PR00010; EGFBL00D.  
DR SMART: SM00032; CCP; 8.  
DR SMART: SM00042; CUB; 3.  
DR SMART: SM00179; EGF\_CA; 8.  
DR SMART: SM00001; EGF\_like; 9.  
DR SMART: SM00231; FA58C; 2.  
DR SMART: SM00282; Lamg; 1.  
DR SMART: SM00192; LDla; 1.  
DR SMART: SM00208; TNFR; 2.  
DR PROSITE: PS00010; ASX\_HYDROXYL; 11.  
DR PROSITE: PS01180; CUB; 6.  
DR PROSITE: PS00022; EGF\_1; 15.  
DR PROSITE: PS01186; EGF\_2; 13.  
DR PROSITE: PS01187; EGF\_CA; 7.  
DR PROSITE: PS01285; FA58C\_1; 1.  
DR PROSITE: PS01209; LDla\_1; 1.  
DR PROSITE: PS00068; LDla\_2; 2.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
SQ SEQUENCE 3396 AA; 369389 MW; E618E9ACEA13E0E5 CRC64;

Best Local Similarity 21.5%; Pred. No. 0.028;  
Matches 51; Conservative 27; Mismatches 96; Indels 63; Gaps 13;  
QY 24 GERALALPEICTOCP-----GSVON-----LSKVAFCYKTRRELMLHARCINOK 69  
DB 3079 GYRILOGVOPSCAACPRLGRTTPKVGASVEECTLPVCSAGTYLNATONMICECR-----K 3133  
QY 70 GTLLDLQNSLEDPGPNFHQHTVYIIDQANPLKDDLANTRGFTQ-----LQT 121  
DB 3134 GYGOSESOQTSLOCP-PN---HSTKITGATS---KSECTNPEHIAEGKPHCDVANNYC 3185  
QY 122 LILPO---HVNCPGGINAMNTITSYIDNOCQGNLCNNTGDPKCPENGSCVPDGP 177  
DB 3186 INVPETSPFKCKCKRGFG-----TGMCTDVCDS-----FCENSGACVKKLG 3229  
QY 178 LLOVCADGFGHYCKMKGSGSFLMFGILGAT--TLVSILLW-----ATORKKAK 227  
DB 3230 TPCSCVCSFTGPHCAERSEFAYIA-GGIAGAVIFIIIVILLWICVSTRKRDPK 3285  
RESULT 9  
QYBLJ1 PRELIMINARY; PRT; 937 AA.  
AC QYBLJ1;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CI-METAL.  
GN CI-METAL.  
OS Clona intestinalis.  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;  
OC Clonidae; Clona.  
OX NCBI\_Taxid=7719;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nakayama A., Satou Y., Satoh N.;  
RT "Isolation and characterization of genes that are expressed during  
RT Clona intestinalis metamorphosis";  
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AB041857; BAB40596.1; -  
SQ SEQUENCE 937 AA; 101043 MW; 8C67830C8E391D07 CRC64;  
Query Match 8.6%; Score 106.5; DB 5; Length 937;  
Best Local Similarity 23.7%; Pred. No. 0.011;  
Matches 41; Conservative 21; Mismatches 66; Indels 45; Gaps 10;  
QY 35 CTGCT--SVONLSKVAE-----YKTTRELMLHARC-----C--NOKGTILGLDQNSL 82  
DB 425 CTNND--CTTCQCKPGYTGNGLVCRDINECLSRSGGVNTNRKRTNTPGSF-----RCVC 478  
QY 83 EDGPNFHQHTVYIIDL---QANPLKGLANTFRGFTQTLTLLPHQVNCPGGINAMNT 139  
DB 479 R---WGYRAGSGRCVDINECKRSPC-GNNARC-----INTGSGTTC-RC 517  
QY 140 ITSVYVVCQCGQKNLCNNTGDPKCPENGSCVPDGPGLLOVCADGFGHYKC 192  
DB 518 NTGYTNGLICRDINECE--AHPNCGENALCI-NELGSRICLCARFGSGPLC 567  
RESULT 10  
QY99940 PRELIMINARY; PRT; 1999 AA.  
AC QY99940;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE NOTCH4.  
GN Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;



DR HSSP: P00740; 1EDM.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_11.  
 DR InterPro: IPR000800; Notch.  
 DR Pfam: PF000023; ank; 6.  
 DR Pfam: PF000008; EGF; 36.  
 DR Pfam: PF000066; notch; 3.  
 DR PRINTS: PR00010; EGFBLDOD.  
 DR PRINTS: PR01452; NOTCH.  
 DR SMART: SM00248; ANK; 4.  
 DR SMART: SM00179; EGF\_CA; 24.  
 DR SMART: SM00001; EGF\_like; 11.  
 DR PROSITE: PS50068; ANK\_REPEAT; 5.  
 DR PROSITE: PS0297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_34.  
 DR PROSITE: PS01186; EGF\_2; 28.  
 DR PROSITE: PS01187; EGF\_CA; 21.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SQ SEQUENCE 2653 AA; 285928 MW; 6AF2A058FECC6329 CRC64;

Query Match 8.9%; Score 111; DB 5; Length 2653;  
 Best Local Similarity 22.2%; Pred. No. 0.012; Mismatches 54; Indels 78; Gaps 11;  
 Matches 44; Conservative 22; Mismatches 54; Indels 78; Gaps 11;

QY 25 VERALALPEICTQCPGVSQVNSKVFAYCK-----TTREIM 59  
 DB 1126 IRRGSLIRQLCINN--GTCKEHGNIHRCYCKQGYTSYCOQELNECESPCLNGTICRL- 1182  
 QY 60 LHAQCLNKGKTIILDLQNCLEDPGPNFQAHTTVIIDLQANPLK-----DLANTFR 114  
 DB 1183 IGSYACVCKR---GFGQNCNELN-----IDDCSPNCGQNGTCHDLVNTF- 1224  
 QY 115 GFTQQLTILPQHVNCPCGINAMNTITSYIDNQICQGGKNCNNTGDEPMCEPSCVDP 174  
 DB 1225 -----SCSCPFG-----TAGLICEVNEVNDCKRGS-----CHNNGTCT-D 1257  
 QY 175 GPGLQVCVADGFHGYKC 192  
 DB 1258 RVGFECACPPGFGVSRG 1275

RESULT 7  
 Q9NGV4 PRELIMINARY; PRT; 1551 AA.  
 AC Q9NGV4;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE SP1070.  
 GN SP1070 OR CG9138.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-ENDOPLASMIC RETICULUM;  
 RA Serano T.L., Pendleton J.D., Rubin G.M.;  
 RT "A reverse genetic screen for genes involved in Drosophila  
 development.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF239608; AAF63500.1;  
 DR FlyBase: FBgn0031879; SP1070.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR001438; EGF\_11.  
 DR InterPro: IPR003410; HYR.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR002172; LDL\_recept\_A.  
 DR Pfam: PF00008; EGF; 16.  
 DR Pfam: PF02494; HYR; 1.  
 DR PRINTS: PR00010; EGFBLDOD.  
 DR PRINTS: PR00011; EGFAMININ.  
 DR SMART: SM00179; EGF\_CA; 6.  
 DR SMART: SM00001; EGF\_like; 9.  
 DR SMART: SM00282; Lmg; 1.  
 DR SMART: SM00192; LDla; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_9.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_15.  
 DR PROSITE: PS01186; EGF\_2; 12.  
 DR PROSITE: PS01187; EGF\_CA; 5.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Repeat.  
 SQ SEQUENCE 1551 AA; 167816 MW; A97EA229E9384F31 CRC64;

Query Match 8.7%; Score 108.5; DB 5; Length 1551;  
 Best Local Similarity 21.5%; Pred. No. 0.012; Mismatches 51; Conservative 27; Mismatches 96; Indels 63; Gaps 13;

QY 24 GVEALALPEICTQCP-----GSVON-----LSKVAFYCKTTRMLHARCLNOK 69  
 DB 1238 GYRLQGVQSPSCACAPGRTPRVGASVECTLPVCSAGTYLINAQNNCIECR-----K 1292  
 QY 70 GTILGLDLQNCLEDPGPNFQAHTTVIIDLQANPLKGDLANFRFPTQ-----LDT 121  
 DB 1293 GYIOSESOQITSLQCP-PN-----HSTKITGATG---KSECTNPCEHIAEKPHCDVNAV 1344  
 QY 122 LILPQ-----HVNCPGGINAMNTITSYIDNQICQGGKNCNNTGDEPMCEPSCVDPG 177  
 DB 1345 IMVPEISDFCECKPFGNG-----TGMACPTVCDS-----FCNSAGCAVCDLKG 1388  
 QY 178 LQQVCVADGFHGYKCMQSGFSILMFFGLIGAT--TLVSILIM-----ATQRRK 227  
 DB 1389 TPSCRCVGFPTGPHCAERSEFAVIA-GGIAGAVIFITIIIVLIMICVSRTRDOK 1444

RESULT 8  
 Q9VW55 PRELIMINARY; PRT; 3396 AA.  
 ID Q9VW55;  
 AC Q9VW55;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CG9138 PROTEIN  
 GN SP1070 OR CG9138.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRATIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

DR PROSITE; PS00070; ALDEHYDE\_DEHYDR\_CYS; UNKNOWN\_1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 43.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE; PS01186; EGF\_2; 36.  
 DR PROSITE; PS01187; EGF\_CA; 43.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 SQ SEQUENCE 2906 AA; 313372 MW; 9EB64E727044EF58 CRC64;

Query Match 9.6%; Score 119.5; DB 11; Length 2906;  
 Best Local Similarity 23.3%; Pred. No. 0.0016;  
 Matches 52; Conservative 17; Mismatches 77; Indels 77; Gaps 14;

OY 1 MAPPGSLTLVPMAMALLALVEEALPEICTO--CPGSONSKAFYCKTREL 58  
 DB 2428 ICPHGPAYD-----GRDIDCKKPSLCTGMLC---VNTGSEFCCKVGYTT 2474  
 OY 59 MLHRCCLNCKGTLGLDLCNCSLEDPGP-NFHOAHTT-----VIIDLQANPLKG 107  
 DB 2475 DISGTAC-----VDLDCS-QSRKPCNFICKNTGSGYSCSPRGVYLDGDKTKC- 2523  
 OY 108 DL-----ANTRGFTQLTLLPHVNCPPGINAMNTTSTYIDNLCGQK 153  
 DB 2524 DLDEQTKQHNCGFLCVNTLGGFT-----CKCPPTGTHH--TACIDNNECGSSO- 2570  
 OY 154 NLCNNTGDPKPCPENGSCVPDGLQVCADGF-----HGKGC 192  
 DB 2571 -----PSLCGAKGIC-QNTPGSFSCGCRGFSLDASGLNC 2604

RESULT 5  
 Q9DE37 PRELIMINARY; PRT: 1515 AA.

ID Q9DE37  
 AC Q9DE37;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE SLIT3.  
 GN SLIT3.  
 OS Brachydanio rerio (zebrafish) (zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.  
 NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MICHIGAN;  
 RA Yeo S.Y., Okamoto H.;  
 RT "zebrafish slit3 homolog";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF210320; AAG36772.1; -  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000359; Cys\_knot.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001791; laminin\_G.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR000483; LRR\_Cterm.  
 DR InterPro; IPR000372; LRR\_Nterm.  
 DR InterPro; IPR003592; LRR\_out.  
 DR InterPro; IPR003591; LRR\_Typ.  
 DR Pfam; PF00008; EGF\_9.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR Pfam; PF00560; LRR; 18.  
 DR Pfam; PF01463; LRRCT; 4.  
 DR Pfam; PF01462; LRRNT; 4.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR SMART; SM00041; CT; 1.  
 DR SMART; SM00181; EGF\_9.  
 DR SMART; SM00179; EGF\_CA; 8.  
 DR SMART; SM00001; EGF\_1like; 6.  
 DR SMART; SM00282; lamc; 1.

DR SMART; SM00370; LRR; 14.  
 DR SMART; SM00082; LRRCT; 4.  
 DR SMART; SM00013; LRRNT; 4.  
 DR SMART; SM00369; LRR\_Typ; 18.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 SQ SEQUENCE 1515 AA; 166871 MW; 12FD111277D18D5B CRC64;

Query Match 9.1%; Score 113; DB 13; Length 1515;  
 Best Local Similarity 22.3%; Pred. No. 0.0039;  
 Matches 59; Conservative 17; Mismatches 69; Indels 120; Gaps 13;

OY 24 GVERALALPEICTCCGSGSONSKAFYC-----KTRRELHARCCLNKG 70  
 DB 704 GAENECCLPS-APRCDESC-TCEIVVRCNSNGLSLPKRAIKDTLXL-----EG 753  
 OY 71 TLGLDLNCSLEDPGPNFHOAHTTVIIDLQANPLKGLANTFNGFTQLTLIL----- 124  
 DB 754 NIL-----TSVPKELPNLKO--LTLLDLSNNSTISHLAPLTFSMWTLATLILSYNOIR 804  
 OY 125 --POHY-----NCPGINAMNTTSTYIDNLCGQK 150  
 DB 805 CVPVHAFDGLRSLRLTLHGNDLSTVPEG-AFNHLTSLSLALGANPLYCD--CELRW 859  
 OY 151 -----GKNMLCNNTGDPK-----CPE 167  
 DB 860 LSGWVKAAGREKRIACCTPPDMADRLTLPLNRCQKGPADLNMKCAPCLATPCQN 919  
 OY 168 NGSCVPDPGGLQVCADGFHGKGC 192  
 DB 920 NGTCVSDVWGSYHCTCPFGYKGRNC 944

RESULT 6  
 Q25253 PRELIMINARY; PRT: 2653 AA.

ID Q25253  
 AC Q25253;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE NOTCH HOMOLOG SCALLOPED WINGS (SCL).  
 GN SCL.  
 OS Lucilia cuprina (greenbottle fly) (Australian sheep blowfly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Cecidoidea; Calliphoridae; Lucilia.  
 NCBI\_TaxID=7375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SS SEEKING;  
 RX MEDLINE=96400928; PubMed=8807304;  
 RA Davies A.G., Game A.Y., Chen Z., Williams T.J., Goodall S., Yen J.L.,  
 RA McKenzie J.A., Batterham P.;  
 RT "Scalloped wings is the Lucilia cuprina Notch homologue and a  
 RT candidate for the modifier of fitness and asymmetry of diazino  
 RT resistance";  
 RL Genetics 143:1321-1337(1996).  
 RN [2]  
 RP SEQUENCE OF 39-265 FROM N.A.  
 RC STRAIN-SS SEEKING;  
 RA Chen Z., McKenzie J.A., Batterham P.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U58977; AAC36151.1; -  
 DR EMBL; AF032672; AAC36152.1; -  
 DR EMBL; AF032670; AAC36152.1; JOINED.  
 DR EMBL; AF032671; AAC36152.1; JOINED.  
 DR EMBL; AF032673; AAC36153.1; -

RESULT 2  
Q9DD14 PRELIMINARY; PRT; 223 AA.  
ID Q9DD14  
AC Q9DD14  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE 0610007C21R1K PROTEIN.  
GN 0610007C21R1K.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Araiwa T., Hata A., Fukushima Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka T.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirral L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Blake J., Boffelli D., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
RA Wyszynski B., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT Functional annotation of a full-length mouse cDNA collection.\*;  
RL Nature 409:685-690(2001).  
DR EMBL: AK002276; BAB21981.1;  
DR MGD: MGI:1918918; 0610007C21R1K.  
DR InterPro: IPR000561; EGF-like.  
DR SMART: SM00181; EGF; 1.  
DR SMART: SM00001; EGF-like; 1.  
DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
SQ SEQUENCE 223 AA; 23867 MW; A85E3A03C7DD2C16 CRC64;

Query Match  
Best Local Similarity 74.8%; Score 929.5; DB 11; Length 223;  
Matches 171; Conservative 19; Mismatches 23; Indels 1; Gaps 1;  
QY 16 AALLLALGVERALALPEICTQCPSSYONLSKVAFCYCTTELMHARRCCLNKGITLGL 75  
DB 11 AALLLVGVERSLALPRITCLTCRGMHLSRVAAYCEDSKSLM-QARRCCCLNKGITLGL 69  
QY 76 DLQNSLEDPCGPNHQHTVYIDLANPLKGDANTFRGTQQTQTLILPOHNCRGGIN 135  
DB 70 NLQNSCLDPCGPNHQHTVYIDLANPLKGDANTFRGTQQTQTLILPOHNCRGGIN 129  
QY 136 AMNTTSTYDNOICOGOKNNTGDEPCPCENKSCVPDGGILQCCVADGFGYKCMRO 195  
DB 130 AMDNVTSTKDKOICOGOKNNTGDEPCPCENKSCVPDGGILQCCVADGFGYKCMRO 189  
QY 196 GSFSLMFFGILGATTLVSLLMATORRRAKTS 229  
DB 190 GSFSLMFFGILGATTLVSLLMATORRRAKTS 223

RESULT 3  
Q9Y5L7 PRELIMINARY; PRT; 208 AA.  
ID Q9Y5L7  
AC Q9Y5L7

DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE APOPTOSIS RELATED PROTEIN APR-3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zhu F., Yan W., Chai Y.B., Shao C., Peng W.D., Yang A.G., Wang C.J.,  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Zhu F., Yan W., Chai Y.B., Shao C., Peng W.D., Yang A.G., Wang C.J.,  
RA Zhao Z.L.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF144055; AAD31317.2;  
DR InterPro: IPR000561; EGF-like.  
DR SMART: SM00001; EGF-like; 1.  
SQ SEQUENCE 208 AA; 22687 MW; CDAC5723C62CA6F CRC64;

Query Match  
Best Local Similarity 63.5%; Score 789; DB 4; Length 208;  
Matches 141; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
QY 47 KVAFCYCTTELMHARRCCLNKGITLGLDQNSLEDPCGPNHQHTVYIDLANPLK 106  
DB 13 KSLYCKTTELMHARRCCLNKGITLGLDQNSLEDPCGPNHQHTVYIDLANPLK 72  
QY 107 GDLANFRGTQQTQTLILPOHNCRGGINAMWNTTSTYDNOICOGOKNNTGDEPCPC 166  
DB 73 GDLANFRGTQQTQTLILPOHNCRGGINAMWNTTSTYDNOICOGOKNNTGDEPCPC 132  
QY 167 ENGSCVPDGGILQCCVADGFGYKCMROG 196  
DB 133 ENGSCVPDGGILQCCVADGFGYKCMROG 162

RESULT 4  
Q9WU9 PRELIMINARY; PRT; 2906 AA.  
ID Q9WU9  
AC Q9WU9;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE FIBRILIN-2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yang Q., Ota K., Tian Y., Kumar A., Wada J., Kashihara N.,  
RA Walling E.I., Kanwar Y.S.;  
RT "Cloning of rat fibrillin-2 cDNA and its role in branching  
RT morphogenesis of rat embryonic lung";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF135060; AAD34439.1;  
DR HSSP: P35555; IEMN.  
DR InterPro: IPR002086; Aldehyde\_dehydr.  
DR InterPro: IPR000152; Asx\_hydroxyl.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001881; EGF Ca.  
DR InterPro: IPR001438; EGF-IT.  
DR InterPro: IPR002212; TB.  
DR Pfam: PF00008; EGF; 46.  
DR Pfam: PF00683; TB; 9.  
DR PRINTS: PR00010; EGFBLDOD.  
DR PRINTS: SM00179; EGF\_CA; 42.  
DR SMART: SM00001; EGF-like; 4.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 08:58:46 ; Search time 135.07 Seconds  
(without alignments)  
247,993 Million cell updates/sec

Title: US-09-726-348-2  
Perfect score: 1243  
Sequence: 1 MAPHGPSTLTVPWMAALL.....TTLSVSLMATORRKAATS 229

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea: \*  
2: SP bacteria: \*  
3: SP fungi: \*  
4: SP human: \*  
5: SP invertebrate: \*  
6: SP mammal: \*  
7: SP mhc: \*  
8: SP organelle: \*  
9: SP phage: \*  
10: SP plant: \*  
11: SP rodent: \*  
12: SP virus: \*  
13: SP vertebrate: \*  
14: SP unclassified: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	940	75.6	171	4 Q9Y2R7	Q9Y2R7 homo sapien
2	929.5	74.8	223	11 Q9PD14	Q9PD14 mus musculu
3	789	63.5	208	4 Q9Y5L7	Q9Y5L7 homo sapien
4	119.5	9.6	2906	11 Q9W0H9	Q9W0H9 rattus norv
5	113	9.1	1515	13 Q9DE37	Q9DE37 brachydanio
6	111	8.9	1551	5 Q9NGV4	Q9NGV4 drosophila
7	108.5	8.7	1551	5 Q9VM55	Q9VM55 drosophila
8	108.5	8.6	1551	5 Q9VM55	Q9VM55 drosophila
9	106.5	8.6	1551	5 Q9VM55	Q9VM55 drosophila
10	106.5	8.6	1551	5 Q9VM55	Q9VM55 drosophila
11	106.5	8.6	1551	5 Q9VM55	Q9VM55 drosophila
12	105.5	8.5	1551	5 Q9VM55	Q9VM55 drosophila
13	105	8.4	2703	5 Q9W4T8	Q9W4T8 drosophila
14	104	8.4	2872	11 Q9W0H8	Q9W0H8 rattus norv
15	104	8.3	585	5 Q9V8E5	Q9V8E5 drosophila
16	103	8.2	1404	5 Q9V8E5	Q9V8E5 drosophila
17	102.5	8.2	2447	13 Q13149	Q13149 fuqu rubrip
18	101.5	8.1	1012	10 Q9LKR24	Q9LKR24 glycine max
19	101	8.1	1012	10 Q9LKR24	Q9LKR24 glycine max

20	101	8.1	1964	11 Q35442	Q35442 mus musculu
21	99.5	8.0	755	11 Q9R0G6	Q9R0G6 mus musculu
22	99	8.0	1064	10 Q9FX19	Q9FX19 arabidopsi
23	98.5	7.9	1584	5 Q93791	Q93791 caenorhabd
24	98	7.9	517	4 Q9NP01	Q9NP01 homo sapien
25	97.5	7.8	708	13 P87363	P87363 gallus gall
26	97.5	7.8	1216	6 Q9BPA0	Q9BPA0 canis fami
27	97.5	7.8	1318	5 Q9VPM0	Q9VPM0 drosophila
28	97.5	7.8	1713	11 Q88349	Q88349 mus musculu
29	96.5	7.8	2471	4 Q04721	Q04721 homo sapien
30	96.5	7.8	2471	4 Q9H240	Q9H240 homo sapien
31	96.5	7.8	2471	4 Q9H240	Q9H240 homo sapien
32	96.5	7.8	2471	4 Q9H240	Q9H240 homo sapien
33	96.5	7.8	2471	4 Q9H240	Q9H240 homo sapien
34	95.5	7.7	1012	10 Q9LKR25	Q9LKR25 glycine max
35	95.5	7.7	1134	10 Q65510	Q65510 arabidopsi
36	94.5	7.6	970	10 Q9M6D8	Q9M6D8 oryza sativ
37	94	7.6	752	13 Q42374	Q42374 brachydanio
38	94	7.6	2352	5 Q61240	Q61240 halocynthia
39	93.5	7.5	619	10 Q9XG61	Q9XG61 sorghum bic
40	93.5	7.5	1003	10 Q49545	Q49545 arabidopsi
41	93.5	7.5	2212	4 Q93072	Q93072 homo sapien
42	93.5	7.5	2570	4 Q9N715	Q9N715 homo sapien
43	93.5	7.5	3312	4 Q9NYQ7	Q9NYQ7 homo sapien
44	93	7.5	1008	10 Q9LKR26	Q9LKR26 glycine max
45	93	7.5	1764	11 Q35806	Q35806 rattus norv

ALIGNMENTS

RESULT 1  
ID Q9Y2R7 PRELIMINARY: PRI: 171 AA.

AC Q9Y2R7; 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE HSPC013.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zhou J., Zhang Q., Fu G., Ye M., Yu Y., Shen Y., Xu S., He K.,  
RA Chen S., Mao M., Chen Z.;  
RT "Human HSPC013 mRNA";  
RL Submitted (15-11-1998) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF077037; AAD27770.1; -  
DR InterPro: I34000561; EGF-like.  
DR SMART; SM00137; EGF\_1;  
DR PROSITE; PS00132; EGF\_1; UNKNOWN\_1.  
DR PROSITE; PS00132; EGF\_2; 1.  
KW EGF-like domain; Glycoprotein.  
SQ  
SEQUENCE 171 AA: 18597 MW: 91A42CD2B2CB0883 CRC64:

Query Match 75.6%; Score 940; DB 4; Length 171;  
Best Local Similarity 100.0%; Pred. No. 5.7e-92;  
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 59 MIHARCCINQKGTILGDLNCSLEDPGPNFHAHTVITIDLANPLKGLANTFEGFTQ 118  
DB 1 MIHARCCINQKGTILGDLNCSLEDPGPNFHAHTVITIDLANPLKGLANTFEGFTQ 60  
OY 119 IOTLIPPHVNCPGINAMNTTTSYIDNOCIGGKNCNTGPEKCPKPGSCVPPGGL 178  
DB 61 IOTLIPPHVNCPGINAMNTTTSYIDNOCIGGKNCNTGPEKCPKPGSCVPPGGL 120  
OY 179 IOTLIPPHVNCPGINAMNTTTSYIDNOCIGGKNCNTGPEKCPKPGSCVPPGGL 229  
DB 121 IOTLIPPHVNCPGINAMNTTTSYIDNOCIGGKNCNTGPEKCPKPGSCVPPGGL 171



Thu Mar 28 09:21:39 2002

us-09-726-348-2\_copy\_126\_177\_1.rspt

---

```

ID 09YGH8 PRELIMINARY: PRT: 1245 AA.
AC 09YGH8;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE INSULIN RECEPTOR PRECURSOR (FRAGMENT).
OS Scaphthalmus maximus (turbot).
OC Euteleostei; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopteleostei; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Pleuronectiformes;
OC Pleuronectidae; Scophthalmidae; Scophthalmus.
OX NCBI_TaxID=52904;
RN [1]
RP SEQUENCE FROM N.A.
RA Elies G., Duval H., Bonnet G., Wolff J., Boeuf G., Boujard D.;
RT "Turbot insulin and insulin-like growth factor-1 receptors: cDNAs
RT cloning and messenger RNAs expression during development.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN
CC TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AJ224994; CAA12279.1; -
DR HSSP: P06213; IIRK.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002011; Receptor_Tyr_Kin_II.
DR InterPro: IPR001245; Tyr_Kin.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 1.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00261; FU; 1.
DR SMART: SM00107; TYRKC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00239; RECEPTOR_TYR_KIN_II; 1.
DR ATP-binding; Glycoprotein; Phosphorylation; Receptor; Transferase;
KW Transmembrane; Tyrosine-protein kinase.
FT NON_TER 1 625
FT CHAIN 1 625 INSULIN RECEPTOR, BETA SUBUNIT.
FT CHAIN 626 1245 INSULIN RECEPTOR, BETA SUBUNIT.
SQ SEQUENCE 1245 AA; 139917 MW; D3DDC5A0C7104C25 CRC64;

```

```

Query Match 20.2%; Score 62.5; DB 13; Length 1245;
Best Local Similarity 29.7%; Pred. No. 17;
Matches 19; Conservative 7; Mismatches 25; Indels 13; Gaps 4;

```

```

OY 1 OHVN-----CPGG---INAMNTIT---SYDNOICOGGK-NLCNNTGDPKPCPENGSCV 47
DB 131 OHGNSCVCKCPAGYVFRGRCVSFSCQELHNOCKKKNKLQNDRESCNEVYIHNGCI 190
OY 48 PDGP 51
DB 191 PEGP 194

```

```

RESULT 15
ID 035442 PRELIMINARY: PRT: 1964 AA.
AC 035442;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NOTCH4.
GN NOTCH4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rower L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF030001; AAB82004.1; -
DR HSSP: P08709; 1BF9.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001438; EGF-II.
DR InterPro: IPR000800; Notch.
DR InterPro: IPR001281; Rleske.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00008; EGF; 27.
DR Pfam: PF00066; notch; 2.
DR PRINTS: PR01415; ANKTRIN.
DR PRINTS: PR00010; EGFBLD.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 5.
DR SMART: SM00179; EGF_CA; 12.
DR SMART: SM00001; EGF_like; 14.
DR PROSITE: PS50088; ANK_REPEAT; 5.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00107; ASX_HYDROXYL; 11.
DR PROSITE: PS00022; EGF_1; UNKNOWN_28.
DR PROSITE: PS01187; EGF_CA; 9.
DR PROSITE: PS00200; RLESKE_2; UNKNOWN_1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 1964 AA; 206699 MW; 53D7F86394FC40BD CRC64;

```

```

Query Match 20.2%; Score 62.5; DB 11; Length 1964;
Best Local Similarity 30.9%; Pred. No. 26;
Matches 17; Conservative 7; Mismatches 18; Indels 13; Gaps 3;
OY 5 CPGGINAMNTITSY-----IDNOICOGGK-NLCNNTGDPKPCPENGSC--VPDG 50
DB 202 CPGGTSCHNTLGSYGQCLCPVGQEGSPQCKLRKGAC----PESCLNGSTQLVPEG 252

```

```

Search completed: March 28, 2002, 09:13:38
Job time: 892 sec

```





RA Pauley A., Stelleyes L.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO PANCREATIC TRYPSTIN INHIBITOR (KUNITZ) DOMAIN.  
 DR EMBL: U29096; AAA68408.1; -  
 DR HSSP: P10646; IPEX.  
 DR WormPep: F30H5.3; CE01927.  
 DR InterPro: IPR002198; ADH\_short.  
 DR InterPro: IPR002899; EB.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR InterPro: IPR002221; WAP.  
 DR Pfam: PF01683; EB; 3.  
 DR Pfam: PF00014; Kunitz\_BPTI; 6.  
 DR PRINTS: PR00759; BASICPTASE.  
 DR SMART: SM00181; EGF; 1.  
 DR SMART: SM00131; KU; 5.  
 DR SMART: SM00217; WAP; 1.  
 DR SMART: SM00289; WRI; 15.  
 DR PROSITE: PS00061; ADH\_SHORT; UNKNOWN\_1.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 2.  
 DR PROSITE: PS00279; BPTI\_KUNITZ\_2; 5.  
 DR Hypothetical protein; Serine protease inhibitor.  
 FT DOMAIN 7  
 FT 1520 1523 POLY-LEU.  
 FT POLY-GLU.  
 SQ SEQUENCE 1599 AA; 171658 MW; AB5E6A1D86E9880D CRC64;

Query Match 21.4%; Score 66; DB 5; Length 1599;  
 Best Local Similarity 27.7%; Pred. No. 7.8;  
 Matches 18; Conservative 6; Mismatches 25; Indels 16; Gaps 2;

OY 2 HVNCPGGINM-----NTITSYDNOIC-----QGSKNLCNNTGDEMCPENGS 45  
 Db 593 YTGCGNDNMFETLLCQNTCENTIPEQCPGQDAYKYGYWYVCSNAGNSCPVAYE 652  
 OY 46 CVPDG 50  
 Db 653 CYFDG 657

RESULT 9  
 ID 09WTS6 PRELIMINARY; PRT; 2715 AA.  
 AC 09WTS6;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE TEN-M3.  
 GN OD23 OR TEN-M3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-BRAIN;  
 RA Ohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T.,  
 RA Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.;  
 RT "Mouse Ten-m/Odz is a new family of dimeric type II transmembrane  
 RT proteins expressed in many tissues.";  
 RL J. Cell Biol. 0:0-0(1999).  
 DR EMBL: AB025412; BAA77398.1; -  
 DR HSSP: P00750; IPEX.  
 DR MGD: MGI:1345183; Odz3.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR Pfam: PF00008; EGF; 7.  
 DR PRINTS: PR00011; EGF\_LAMININ.  
 DR SMART: SM00181; EGF; 6.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_8.  
 DR PROSITE: PS01186; EGF\_2; 7.  
 KW EGF-like domain; Glycoprotein.  
 SQ SEQUENCE 2715 AA; 303063 MW; 598FA6A77334C2E1 CRC64;

Query Match 21.28; Score 65.5; DB 11; Length 2715;  
 Best Local Similarity 30.0%; Pred. No. 15;  
 Matches 18; Conservative 5; Mismatches 20; Indels 17; Gaps 3;

OY 1 QHVNCPGGINM-----NTITSYDNOICQGSKNLCNNTGDEMCPENGSCVDPDG 52  
 Db 718 EHGCKDKCKSCGQWNGEHTIHYDKIVKESCPGLNS-----NGRCTLIDNG 768

RESULT 10

ID 09RUC8 PRELIMINARY; PRT; 747 AA.

AC 09RUC8;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE SECRETED ENDOGLUCANASE.  
 GN CELL.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Oliver K., Harris D.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmid and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL131171; CAB61539.1; -  
 DR HSSP: P14090; IULO.  
 DR InterPro: IPR003305; CBD\_6.  
 DR InterPro: IPR001701; Glyco\_hydro\_9.  
 DR InterPro: IPR000566; Lipocin\_cytfABP.  
 DR Pfam: PF02018; CBD\_6; 1.  
 DR Pfam: PF00759; Glyco\_hydro\_9; 1.  
 DR PROSITE: PS00698; GLYCOSYL\_HYDROL\_F9\_2; 1.  
 DR PROSITE: PS00213; LIPOCALIN; UNKNOWN\_1.  
 SQ SEQUENCE 747 AA; 80493 MW; 72D33C07452B7716 CRC64;

Query Match 20.9%; Score 64.5; DB 2; Length 747;  
 Best Local Similarity 30.0%; Pred. No. 5.7;  
 Matches 15; Conservative 10; Mismatches 10; Indels 15; Gaps 2;

OY 7 GGINAMNTITSYDNOICQGSKNLCNNTGPEM-----CPENGSCVDP 49  
 Db 351 GGISTWELTSTY-----ERSLHARTGPKLGDGSDIDPESGNRPD 392

RESULT 11

ID 022378 PRELIMINARY; PRT; 1101 AA.

AC 022378;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE CODED FOR BY C. ELEGANS CDNA YK23A6.3.  
 GN T10E10.4.

RN	SEQUENCE FROM N.A.
RP	TISSUE=DORSAL ROOT GANGLION;
RC	MEDLINE=98024190; Pubmed=9356508;
RA	Friedel R.H., Schuercher H., Strubusch J., Barde Y.A.;
RT	"Identification of genes differentially expressed by nerve growth
R7	factor-and neurotrophin-3-dependent sensory neurons.";
RL	Proc. Natl. Acad. Sci. U.S.A. 94:12670-12675(1997).
DR	EMBL: AJ011489; CAA04784.1; "-
DR	InterPro: IPR000636; Channel_chan_non_119.
DR	InterPro: IPR001682; Channel_pore_Ca_Na.
DR	Pfam: PF00520; ion_trans; 1.
KW	Ionic channel.
FT	NON_TER 1
FT	NON_TER 213 213
SO	SEQUENCE 213 AA; 24397 MW; DAF3B6AE4E8E47ED CRC64;

Query Match	22.3%	Score 69;	DB 13;	Length 213;	
Best Local Similarity	40.0%;	Pred No. 0.45;			
Matches 16;	Conservative	6;	Mismatches 12;	Indels 6;	Gaps 2

OY	17 SYDNOI-----CQGQKN--LCNNTPGPEMCPENGSCVPDG 50				
	: :: :	:   :	:: :	:: :	:: :
D6	110 AITENSHFYRLREGOKVYLGGNSDAGKCFEEFICVKAG 149				

RESULT	6
Q9PVZ4	PRELIMINARY; PRT; 1362 AA.
ID	09PVZ4
AC	09PVZ4;
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	INSULIN RECEPTOR PRECURSOR.
OS	Xenopus laevis (African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC	Xenopodinae; Xenopus.
OX	NCBI_TaxId=8335;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Groino L., Richard-Parpallion L., Boujard D.;
RT	"Expression pattern of insulin receptor mRNA during Xenopus laevis
RT	embryogenesis.";
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC	-1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN
CC	TYROSINE PHOSPHATE.
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC	HSP: P06213; IIRK.
DR	EMBL: AJ132556; CAB46565.1; "-
DR	HSP: P06213; IIRK.
DR	InterPro: IPR000494; EGFR.L.
DR	InterPro: IPR000719; Euk_ptkinase.
DR	InterPro: IPR003961; FN_III.
DR	InterPro: IPR002174; Furin-like.
DR	InterPro: IPR002011; Receptor Tyr_kin_II.
DR	InterPro: IPR001245; Tyr_kin.
DR	Pfam: PF000641; fn3; 1.
DR	Pfam: PF00757; Furin-like; 1.
DR	Pfam: PF00069; Pkinase; 1.
DR	Pfam: PF01030; Recep_dom; 2.
DR	PRINTS: PR00109; TYRKINASE.
DR	SMART: SMO0060; FN3; 2.
DR	SMART: SMO0261; FU; 2.
DR	SMART: SMO0219; TYRKC; 1.
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
DR	PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR	PROSITE: PS00239; RECEPTOR_TYR_KIN_II; 1.
KW	ATP-binding; Glycoprotein; Phosphorylation; Receptor; Signal;
KW	Transferase; Transmembrane; Tyrosine-Protein kinase.
FT	SIGNAL 1 37
FT	CHAIN 38 754 INSULIN RECEPTOR, ALPHA-SUBUNIT.

```

FT CHAIN 759 1362 INSULIN RECEPTOR, BETA-SUBUNIT.
SO SEQUENCE 1362 AA; 153755 MW; 7B8BF2F87EFD0A1B CRC64;

Query Match 21.7%; Score 67; DB 13; Length 1362;
Best local similarity 33.3%; Pred. No. 5;
Matches 18; Conservative 3; Mismatches 21; Indels 12; Gaps 4;

OY 5 CPGG---INAMNTTSTYIDNLCGQKNCNNCGDPMCP----ENMGSCVDPGP 51
||| ||| ||| : | : | : | : | : | : | : | : | : | : | : | : |
DB 278 CPGSGYOFQGWRG-----IDFNTCOELNSRCQNSKD-NSCEPYIHKGECHPDCP 326

RESULT 7
O9JUV9 PRELIMINARY; PRT: 2019 AA.
ID O9JUV9
AC O9JUV9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE VOLTAGE-GATED SODIUM CHANNEL.
GN SCNSA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=HEART;
RA Zimmer T., Benddorf K.;
RT "The mouse heart sodium channel (mhl): cloning and characterization of
alternatively spliced variants."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
EMBL AJ271477; CAB70096.1; -.
MD MG1:98251; Scn5a.
DR InterPro: IPR000636; Cation_chan_non_lig.
DR InterPro: IPR002111; Cat_channel_trpL.
DR InterPro: IPR001682; Channel_pore_Ca_Na.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR001696; Na_channel.
DR Pfam: PF00520; ion_trans_4.
DR Pfam: PF00612; IQ_1.
DR PRINTS: PR00170; NACHANNEL.
DR SMART: SM00015; IQ_1.
KW Ionic channel.
KW SEQUENCE 2019 AA; 227620 MW; FDB48750D599B89A CRC64;

Query Match 21.7%; Score 67; DB 11; Length 2019;
Best local similarity 29.6%; Pred. No. 7.4;
Matches 16; Conservative 8; Mismatches 20; Indels 10; Gaps 2;

OY 7 GGNA----WNTITSYIDNLCGQKNCNNCGDPMCP-----LCNNTGDPEMCPENGSCVPDG 50
||| ||| ||| : | : | : | : | : | : | : | : | : | : | : | : |
DB 232 GSVEADGIWNSLDYLNDRPNAYLLKNGTDDVLICGSSDAGTCPEGYRCLAKG 345

RESULT 8
O09983 PRELIMINARY; PRT: 1599 AA.
ID O09983
AC O09983;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 171.7 KDA PROTEIN F30H5.3 IN CHROMOSOME III.
F30H5.3.
GN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;

```

01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE APOPTOSIS RELATED PROTEIN APR-3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zhao Z.L.;  
 RA Zhao Z.L.;  
 RA Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Zhu F.; Yan W.; Chai Y.B.; Shao C.; Peng W.D.; Yang A.G.; Wang C.J.;  
 RA Zhao Z.L.;  
 RA Submitted (Jun-1999) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AF144055; AAD31317.2; -  
 DR InterPro: IPR000561; EGF-like.  
 DR SMART: SM00001; EGF-like; 1.  
 SO SEQUENCE 208 AA; 22687 MW; CD4C5723C62CAF6F CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 309; DB 4; Length 208;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

01 0HVNCGGINAMNNTTSTYDNOICOGKNCNNTGDEPCMGSCVPDGP 52  
 DB 92 OHVNCGGINAMNNTTSTYDNOICOGKNCNNTGDEPCMGSCVPDGP 143

RESULT 3  
 ID 09DD14 PRELIMINARY; PRT: 223 AA.

DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE 0610007C21RIK  
 GN 0610007C21RIK  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;  
 RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J.; Shinagawa A.; Shibata K.; Yoshino M.; Itoh M.; Ishii Y.;  
 RA Aizawa K.; Izawa M.; Nishi K.; Kiyosawa H.; Kondo S.; Yamanaka I.;  
 RA Saito T.; Okada Y.; Gojohori T.; Bono H.; Kasukawa T.; Saito R.;  
 RA Kadota K.; Matsuda H.A.; Ashburner M.; Batilov S.; Casavant T.;  
 RA Fleischmann W.; Gaasterland T.; Gissi C.; King B.; Kocha H.;  
 RA Kuehl P.; Lewis S.; Matsuo Y.; Nikaido I.; Resole G.; Quackenbush J.;  
 RA Kuehl P.; Staudt F.; Suzuki R.; Tomita M.; Wagner L.; Washio T.;  
 RA Sakai K.; Okido T.; Furuno M.; Bono H.; Badarelli R.; Barsh G.;  
 RA Blake J.; Boffelli D.; Bojunga N.; Carninci P.; de Bonaldo M.F.;  
 RA Brownstein M.J.; Bull C.; Fletcher C.; Fujita M.; Gariboldi M.;  
 RA Guastacchi S.; Hill D.; Hermann M.; Hume D.A.; Kamiya M.; Lee N.H.;  
 RA Lyons P.; Matchonni L.; Mashima J.; Mazzarelli J.; Mombauts P.;  
 RA Nordone P.; Ring B.; Ringwald M.; Rodriquez I.; Sakamoto N.;  
 RA Sasaki H.; Sato K.; Schoenbach C.; Seya T.; Shibata Y.; Storch K.-F.;  
 RA Suzuki H.; Toyooka K.; Wang K.H.; Weitz C.; Whitaker C.; Wilming L.;  
 RA Wyshaw-Boris A.; Yoshida K.; Hasegawa Y.; Kawai H.; Kohlsuki S.;  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:665-690(2001).  
 DR EMBL: AK002276; BAB21981.1; -  
 DR MGD: MGI:1918918; 0610007C21RIK.  
 DR InterPro: IPR000561; EGF-like.  
 DR SMART: SM00001; EGF-like; 1.  
 DR SMART: SM00001; EGF-like; 1.

DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 SO SEQUENCE 223 AA; 23867 MW; A85E3A03C7DD2C16 CRC64;

Query Match  
 Best Local Similarity 71.2%; Score 227; DB 11; Length 223;  
 Matches 37; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

01 0HVNCGGINAMNNTTSTYDNOICOGKNCNNTGDEPCMGSCVPDGP 52  
 DB 120 OHVNCGGINAMNNTTSTYDNOICOGKNCNNTGDEPCMGSCVPDGP 171

RESULT 4

ID 096687 PRELIMINARY; PRT: 242 AA.

DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE FIBRINOLYTIC ENZYME (FRAGMENT).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lumbriata; Mollusca; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
 OC Lumbriata; Mollusca; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
 NCBI\_TaxID=86417;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Liang G.; Meng X.; Fu S.; Li L.; Hou Y.; Chen F.; Shan F.; Qin Y.;  
 RA "Cloning of the genes encoding fibrinolytic enzymes from Lumbricus  
 RL Submitted (Nov-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
 CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
 DR EMBL: AF109648; AAD05563.1; -  
 DR HSSP: P00750; 1RTP.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP-SPE; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 KW Hydrolyase; Serine protease.  
 FT NON\_TER 1  
 FT TER 1  
 SO SEQUENCE 242 AA; 24837 MW; F912425D2724745D CRC64;

Query Match  
 Best Local Similarity 24.9%; Score 77; DB 5; Length 242;  
 Matches 20; Conservative 5; Mismatches 13; Indels 28; Gaps 4;

01 7 GGINAMNNTTSTYDNOICOGKNCNNTGDEPCMGSCVPDGP 45  
 DB 166 GGANW-----DNHICVDPAGNTACNDSGGPLNCPDGTGVVGVSVVSGSLGT 218  
 QY 46 CVPDGP 51  
 DB 219 CVPDGP 224

RESULT 5

ID 042419 PRELIMINARY; PRT: 213 AA.

DT 01-JAN-1998 (Tremblrel. 05, Created)  
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
 DE VOLTAGE-GATED SODIUM CHANNEL I (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 NCBI\_TaxID=9031;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2002, 09:13:35 ; Search time 135.07 Seconds  
(without alignments)  
56.313 Million cell updates/sec

Title: US-09-726-348-2\_COPY\_126\_177  
Sequence: 1 QHVNCPGGINAMWTITSYD.....NNGDEPMCEPGSCVDPDGP 52

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP archaea:\*  
2: SP bacteria:\*  
3: SP fungi:\*  
4: SP human:\*  
5: SP invertebrate:\*  
6: SP mammal:\*  
7: SP mhc:\*  
8: SP organelle:\*  
9: SP phage:\*  
10: SP plant:\*  
11: SP protist:\*  
12: SP virus:\*  
13: SP vertebrate:\*  
14: SP unclassified:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	309	100.0	171	4 Q9Y2R7	Q9Y2R7 homo sapien
2	309	100.0	208	4 Q9Y5L7	Q9Y5L7 homo sapien
3	227	73.5	223	11 Q9BD14	Q9BD14 mus musculu
4	77	24.9	242	5 Q9687	Q9687 lumbricus b
5	69	22.3	213	13 Q42419	Q42419 gallus gall
6	67	21.7	1362	13 Q9PVZ4	Q9PVZ4 xenopus lae
7	67	21.7	2019	11 Q9JIV9	Q9JIV9 mus musculu
8	66	21.4	1599	5 Q09983	Q09983 caenorhabdi
9	65.5	21.2	2715	11 Q9WTS6	Q9WTS6 mus musculu
10	64.5	20.9	747	2 Q9R3C8	Q9R3C8 streptomyce
11	64.5	20.9	1101	5 Q22378	Q22378 caenorhabdi
12	64.5	20.9	1841	11 Q9ER60	Q9ER60 mus musculu
13	63.5	20.6	225	13 Q42420	Q42420 gallus gall
14	62.5	20.2	1245	13 Q9YGH8	Q9YGH8 scophthalmu
15	62.5	20.2	1964	11 Q35442	Q35442 mus musculu
16	62.5	20.2	2471	4 Q04721	Q04721 homo sapien
17	62	20.1	226	12 Q9DIX1	Q9DIX1 hepatitis b
18	62	20.1	247	13 Q9IBD6	Q9IBD6 takifugu pa
19	62	20.1	300	12 Q9YPC9	Q9YPC9 chimpanzee

20	62	20.1	2872	11 Q9WUH8	Q9WUH8 ratius norv
21	61.5	19.9	161	6 Q9GK98	Q9GK98 pan troglod
22	61.5	19.9	224	4 Q9BGP3	Q9BGP3 homo sapien
23	61.5	19.9	1834	6 Q28371	Q28371 equus cabal
24	61	19.7	358	11 P97946	P97946 mus musculu
25	61	19.7	441	5 Q9W5X1	Q9W5X1 dirosophila
26	61	19.7	1905	5 Q9XTP6	Q9XTP6 plasmodium
27	61	19.7	2906	11 Q9WUH9	Q9WUH9 ratius norv
28	61	19.7	4599	11 Q9JTI8	Q9JTI8 mus musculu
29	60.5	19.6	497	12 Q69288	Q69288 turkey herp
30	60.5	19.6	1965	5 Q61893	Q61893 caenorhabdi
31	60.5	19.6	2005	4 Q9BZD0	Q9BZD0 homo sapien
32	60.5	19.6	2005	4 Q9BZC9	Q9BZC9 homo sapien
33	60.5	19.6	2771	11 Q9WTS7	Q9WTS7 mus musculu
34	60	19.4	211	13 Q42418	Q42418 gallus gall
35	60	19.4	1194	13 Q9W737	Q9W737 gallus gall
36	60	19.4	1880	13 Q9IBF1	Q9IBF1 takifugu pa
37	60	19.4	1951	4 Q9C007	Q9C007 homo sapien
38	60	19.4	1951	4 Q9BZB3	Q9BZB3 homo sapien
39	60	19.4	2000	4 Q9C006	Q9C006 homo sapien
40	60	19.4	2447	13 Q13149	Q13149 fuigu rubrip
41	60	19.4	3857	11 Q88840	Q88840 mus musculu
42	60	19.4	4599	4 Q9NZR2	Q9NZR2 homo sapien
43	59.5	19.3	296	2 Q9YED3	Q9YED3 escherichia
44	59.5	19.3	389	5 Q21081	Q21081 caenorhabdi
45	59.5	19.3	495	2 Q66606	Q66606 aquifex aeo

## ALIGNMENTS

RESULT 1  
ID Q9Y2R7 PRELIMINARY; PRT; 171 AA.  
AC Q9Y2R7;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE HSPC013.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
ON NCBI\_TaxID=9606;  
RX [1]  
RP SEQUENCE FROM N.A. Zhou J., Zhang Q., Fu G., Ye M., Yu Y., Shen Y., Xu S., He K.,  
RA Chen S., Mao M., Chen Z.;  
RT "Human HSPC013 mRNA";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF077037; AAD2770.1; "  
DR InterPro: IPR00561; EGF-like.  
DR SMART: SMC\_001; EGF\_1.  
DR PROSITE: PS03022; EGF\_1; UNKNOWN\_1.  
DR PROSITE: PS01186; EGF\_2; 1.  
KW EGF-like domain; Glycoprotein.  
SQ  
SEQUENCE 171 AA; 18597 MW; 91A42CDB2CB0883 CRC64;

Query Match 100.0%; Score 309; DB 4; Length 171;  
Best Local Similarity 100.0%; Pred. No. 2.3e-31;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QHVNCPGGINAMWTITSYDNOICGQKNCNNNGDEPMCEPGSCVDPDGP 52  
DB 68 QHVNCPGGINAMWTITSYDNOICGQKNCNNNGDEPMCEPGSCVDPDGP 119  
RESULT 2  
ID Q9Y5L7 PRELIMINARY; PRT; 208 AA.  
AC Q9Y5L7;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

1  
2  
3  
4

Thu Mar 28 09:21:31 2002

us-09-726-348-2.rpr

A:Accession: S10850  
 A:Molecule type: Protein  
 A:Residues: 1-25 <ROB>  
 C:Superfamily: wheat alpha-amylase inhibitor  
 C:Keywords: alpha-amylase inhibitor

Query Match 2.2%; Score 5; DB 2; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 169 GSCVP 173  
 |||||  
 DB 2 GSCVP 6

RESULT 14  
 A32806  
 transposase - Escherichia coli transposon Tn4651 (fragment)  
 C:Species: Escherichia coli  
 C:Date: 20-Oct-1989 #sequence\_revision 20-Oct-1989 #text\_change 20-Jun-2000  
 C:Accession: A32806  
 R:Tsuda, M.; Minegishi, K.I.; Iino, T.  
 J. Bacteriol. 171, 1386-1393, 1989  
 A:Title: Toluene transposons Tn4651 and Tn4653 are class II transposons.  
 A:Reference number: A32806; M01D:89155438  
 A:Accession: A32806  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-26 <TSU>  
 A:Cross-references: GB:D90148; NID:g217162; PIDN:BA14178.1; PID:g217164  
 C:Superfamily: transposase Tn21

Query Match 2.2%; Score 5; DB 2; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 29 LAUPE 33  
 |||||  
 DB 16 LAUPE 20

RESULT 15  
 B44107  
 cytochrome P450 NF3, beta-naphthoflavone-induced (N-terminal) - chicken (fragment)  
 N:Contains: oxidoreductase (EC 1.-.-.-)  
 C:Species: Gallus gallus (chicken)  
 C:Date: 27-Apr-1993 #sequence\_revision 20-Aug-1994 #text\_change 05-Mar-1999  
 C:Accession: B44107  
 R:Nakai, K.; Ward, A.M.; Gannon, M.; Rifkind, A.B.  
 J. Biol. Chem. 267, 19503-19512, 1992  
 A:Title: Beta-naphthoflavone induction of a cytochrome P-450 arachidonic acid epoxigenase  
 A:Reference number: A44107; M01D:92406903  
 A:Accession: B44107  
 A:Molecule type: protein  
 A:Residues: 1-26 <NAK>  
 A:Experimental source: embryo liver  
 A:Note: sequence extracted from NCBI backbone (NCBIP:113919)  
 C:Superfamily: unassigned cytochrome P450; cytochrome P450 homology  
 C:Keywords: electron transfer; heme; monooxygenase; oxidoreductase

Query Match 2.2%; Score 5; DB 2; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 17 AALLL 21  
 |||||  
 DB 5 AALLL 9

Search completed: March 28, 2002, 09:17:27  
 Job time: 381 sec

## RESULT 8

B30609  
Ig kappa chain V-III region (She) - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 29-Jun-1989 #sequence\_revision 29-Jun-1989 #text\_change 30-May-1997  
C:Accession: B30609  
R:Goni, F.R.; Chen, P.P.; McGinnis, D.; Arjonilla, M.L.; Fernandez, J.; Carson, D.; Sold  
J. Immunol. 142, 3158-3163, 1989  
A:Title: Structural and idiotypic characterization of the L chains of human IgM autoanti  
A:Reference number: A30601; MUID:89215279  
A:Accession: B30609  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-24 <GON>  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
C:Keywords: heterotrimer; immunoglobulin

Query Match 2.2%; Score 5; DB 2; Length 24;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 211 TLVS 215  
DB 10 TLVS 14

## RESULT 9

PC0677  
Photosystem I 9.0K H2 chain - common tobacco (fragment)  
C:Species: Nicotiana tabacum (common tobacco)  
C>Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 17-Mar-1999  
C:Accession: PQ0677  
R:Obokata, J.; Mikami, K.; Hayashida, N.; Nakamura, M.; Sugiyama, M.  
Plant Physiol. 102, 1259-1267, 1993  
A:Title: Molecular heterogeneity of photosystem I. psad, psae, psaf, psah and psal are  
A:Reference number: PQ0677; MUID:94105345  
A:Accession: PQ0677  
A:Molecule type: protein  
A:Residues: 1-24 <GON>  
C:Superfamily: photosystem I protein psah  
C:Keywords: chloroplast; photosynthesis; photosystem I; thylakoid

Query Match 2.2%; Score 5; DB 2; Length 24;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 DLANT 112  
DB 13 DLANT 17

## RESULT 10

S10681  
probable 7-ethoxycoumarin O-deethylase (EC 1.14.14.-) cytochrome P450 isoform 3 - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 05-Mar-1999  
C:Accession: S10681  
R:Sincalir, J.F.; Wood, S.; Lambrecht, L.; Gorman, N.; Mende-Mueller, L.; Smith, L.; Hun  
Biochem. J. 269, 85-91, 1990  
A:Title: Isolation of four forms of acetone-induced cytochrome P-450 in chicken liver by  
A:Reference number: S10680; MUID:90328998  
A:Accession: S10681  
A:Molecule type: protein  
A:Residues: 1-24 <STN>  
C:Superfamily: unassigned cytochrome P450; cytochrome P450 homology  
C:Keywords: electron transfer; endoplasmic reticulum; heme; membrane protein; monooxygen

Query Match 2.2%; Score 5; DB 2; Length 24;

Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AALL 21  
DB 5 AALL 9

## RESULT 11

S10682  
cytochrome P450 isoform 4 - chicken (fragment)  
N:Contains: oxidoreductase (EC 1.-.-.-)  
C:Species: Gallus gallus (chicken)  
C>Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 05-Mar-1999  
R:Sincalir, J.F.; Wood, S.; Lambrecht, L.; Gorman, N.; Mende-Mueller, L.; Smith, L.;  
Biochem. J. 269, 85-91, 1990  
A:Title: Isolation of four forms of acetone-induced cytochrome P-450 in chicken liver  
A:Reference number: S10680; MUID:90328998  
A:Accession: S10682  
A:Molecule type: protein  
A:Residues: 1-24 <STN>  
C:Superfamily: unassigned cytochrome P450; cytochrome P450 homology  
C:Keywords: electron transfer; endoplasmic reticulum; heme; membrane protein; monooxy

Query Match 2.2%; Score 5; DB 2; Length 24;  
Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 17 AALL 21  
DB 5 AALL 9

## RESULT 12

PC1221  
alanine racemase (EC 5.1.1.1) - Pseudomonas fluorescens (fragment)  
C:Species: Pseudomonas fluorescens  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999  
C:Accession: PC1221  
R:Yokoiyawa, K.; Kawal, H.; Endo, K.; Iim, Y.H.; Esaki, N.; Soda, K.  
Biosci. Biotechnol. Biochem. 57, 93-97, 1993  
A:Title: Thermolabile alanine racemase from a psychrotroph, Pseudomonas fluorescens:  
A:Reference number: PC1221; MUID:93169026  
A:Accession: PC1221  
A:Molecule type: protein  
A:Residues: 1-25 <YOK>  
C:Comment: This enzyme catalyzes the racemization of L- and D-alanine.  
C:Superfamily: alanine racemase  
C:Keywords: isomerase; pyridoxal phosphate

Query Match 2.2%; Score 5; DB 2; Length 25;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 IDLQA 302  
DB 8 IDLQA 12

## RESULT 13

S10850  
alpha-amylase inhibitor - durum wheat (fragment)  
N:Alternate names: glutenin low molecular weight chain  
C:Species: Triticum durum (durum wheat)  
C>Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 29-Aug-1997  
C:Accession: S10850  
R:Kobrehel, K.; Alary, R.  
J. Sci. Food Agric. 48, 441-452, 1989  
A:Title: Isolation and partial characterisation of two low molecular weight durum whe  
A:Reference number: S10849